INFORMATION COUPLING
perceptions and reality

Rastko Vuković

Banja Luka, 2024.
From Serbian, 2025.

Sadržaj

Glava 1

Introduction

This book is a companion to the blog that I ran from June 2021 to the end of 2024, intended to provide simpler answers to some questions of interest to my information theory.

1.1 Conservation Law

We know energy is subject to the *conservation law*. Energy can change form—potential, thermal, or kinetic—but not its total amount. An isolated, or *closed system*, is a larger framework in which the total energy remains constant. The law of conservation of energy states that energy cannot be created or destroyed; it can only be transferred from one form to another.

For example, water can generate electricity. When water falls from the sky, it converts potential energy into kinetic energy. This energy is then used to rotate a turbine in a generator to produce electricity. The potential energy of water in a hydroelectric dam can be changed into kinetic energy. Then, light bulbs can transform electrical energy into light, which illuminates dark spaces.

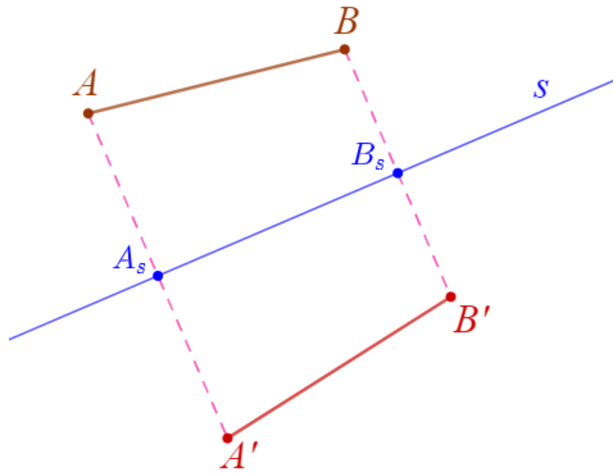
In the remainder of the text, I repeat some well-known phenomena so that we can remember them and establish the expressions we will see with them.

1.1.1 Light

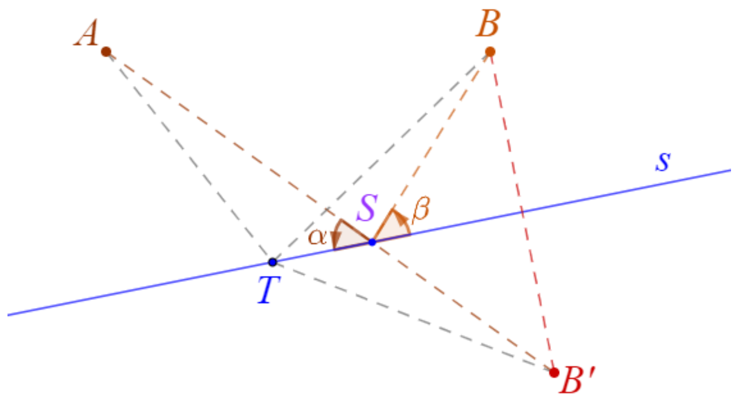
Isometry is any mapping that preserves the distance between points. In geometry, these are translation, axial symmetry, mirror symmetry, and rotation. As we know, they can all be reduced to rotations.

Example 1. *In the image of axial symmetry, let's show that it is an isometry.*

Proof. In the figure, the axis of symmetry $\rho_s : AB \to A'B'$ is the line AB in the line $A'B'$ about the axis s. Here, $AA' \perp s$, $BB' \perp s$, $AA_s = A_sA'$, $BB_s = B_sB'$, and, of course, $AB = A'B'$. Here, $A_s = AA' \cap s$ and $B_s = BB' \cap s$ are the intersection points of the joints of the original and the image with the axis of symmetry, the line s. □



Example 2. *In the image of reflected light, the light path is the shortest.*



Proof. We imagine that a ray of light from point A goes to point B after *reflection* from a mirror, axis s at point S. It behaves as if it were blindly going towards the reflection B', so when it hits the mirror at point S, it continues towards point B.

That $A - S - B$ is the shortest path of light can also be seen using isometry. Namely, the axisymmetric mapping is $\rho_s : B \to B'$, so $SB = SB'$ and $\angle sSB = \angle B'Ss$, and hence $\angle AST = \angle sSB$, so $A - S - B'$ is a straight line. Therefore $AB' < AT + TB'$ (triangle inequality), and since $TB = TB'$ then $AS + SB \le AT + TB'$ for an arbitrary point $T \in s$. $\square$

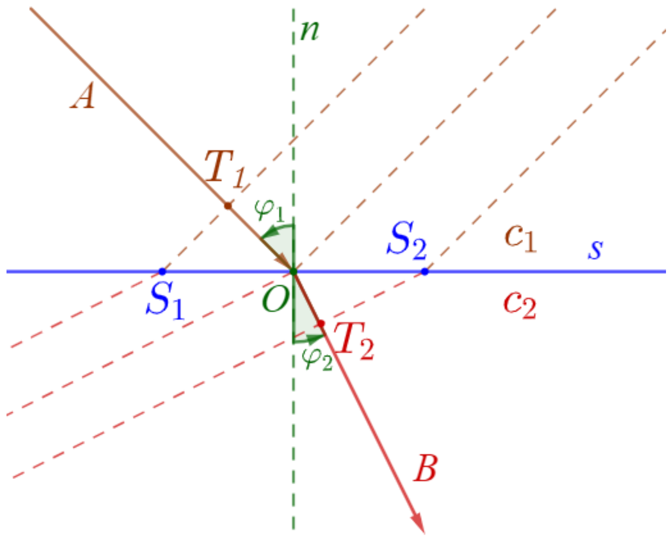
Indeed, light is reflected in such a way that it travels the shortest path, in the shortest time, at a constant speed c. The angle of incidence of light is equal to the angle of reflection in the figure $\alpha = \angle AST = \angle sSB = \beta$. Moreover, by adhering to this minimum, light follows the conservation law that is actually encompassed by the isometry itself.

Snell's law talks about the *refraction* of light that passes into another optical density, from a medium in which it had a speed of c_1 to a medium in which it has a speed of c_2. In the following figure, $c_2 < c_1$, so the angle of the incoming light ray AO to the normal n to the boundary of the medium s is marked φ_1. In the lower middle is the outgoing light ray OB with an angle to the normal φ_2. Two proofs of *Snell's law* can be found in my attachment Speed of light, one of which I will now cite as an example.

Example 3. *For the incident and exit angles, φ_1 and φ_2, with respect to the normal n to the boundary s of the mean of the two optical densities of the light wave speeds c_1 and c_2, the following equality holds:*

$$\frac{\sin\varphi_1}{\sin\varphi_2} = \frac{c_1}{c_2}.$$

This is Snell's law.



Proof. In the figure, the dashed lines perpendicular to the direction of light are wavefronts. The spaces between them are the wavelengths, the upper $\lambda_1 = c_1\Delta t$ and the lower $\lambda_2 = c_2\Delta t$, which the light travels in equal time intervals Δt. Then the intervals $S_1O = OS_2 = d$ are also equal, as are the angles with the perpendicular arms $\angle OS_1T_1 = \varphi_1$, or $\angle OS_2T_2 = \varphi_2$. Therefore, $T_1O = d\cdot\sin\varphi_1$ and $OT_2 = d\cdot\sin\varphi_2$, or:

$$\lambda_1 = d\cdot\sin\varphi_1, \quad \lambda_2 = d\cdot\sin\varphi_2,$$

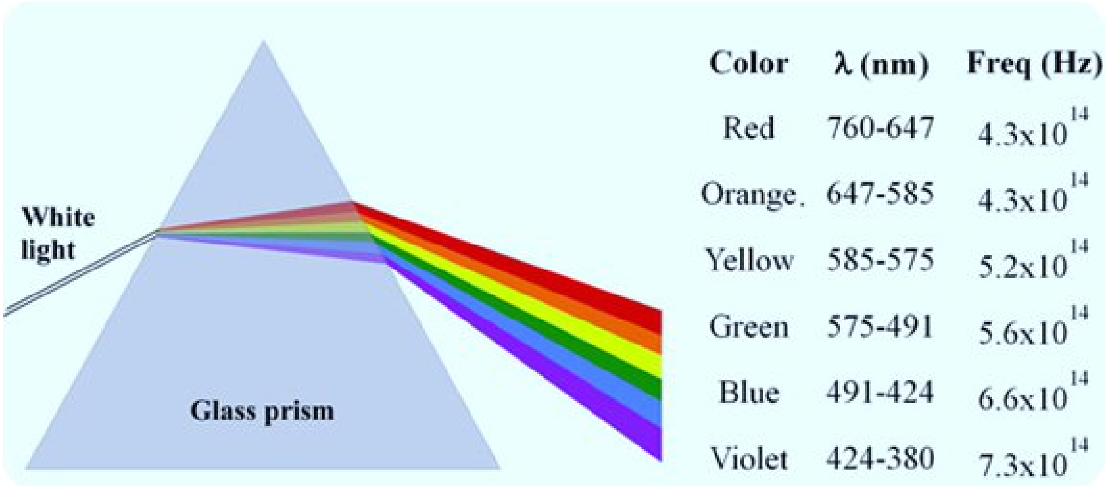
$$c_1\Delta t = d\cdot\sin\varphi_1, \quad c_2\Delta t = d\cdot\sin\varphi_2,$$

$$\frac{c_1\Delta t}{\sin\varphi_1} = \frac{c_2\Delta t}{\sin\varphi_2},$$

$$\frac{\sin\varphi_1}{\sin\varphi_2} = \frac{c_1}{c_2},$$

which proves the aforementioned Snell's law. □

Example 4. *In the following image, white sunlight passes through a glass prism and is split into six*[1] *colors.*



Violet light has shorter wavelengths ($\lambda = 380 - 450$ nm) than red light ($\lambda = 625 - 750$ nm) and, as can be seen in the image of a glass prism, bends it more downward. A glass *prism* is more refractive because shorter wavelengths "stick" to the particles of that optical medium more, interacting more, that is, communicating with the environment. Inside a glass prism, red light travels the fastest of the colors *white light*, because it has the lowest frequency and the longest wavelength.

Leaving the prism, the light continues at the first speed but is split into separate wavelengths (400–700 nm). White light is thus VIBGYOR with energies $E = hf$, where $h = 6.62607015{\times}10^{-34}$ Js *Planck's constant* and f in the figure with the frequencies indicated, colored components. The question of wavelength and frequency ($\lambda f = c$) thus becomes a question of average values to which statistics respond in different ways.

1.1.2 Hartley's Information

Hartley (1928) defined information $H = \log_b N$ as the logarithm of the number $N \in \mathbb{N}$ of equally likely outcomes. When the base of the logarithm $b = 2$, we are dealing with *binary* information that we will understand by tossing a (fair) *coin*. The information carried by such an outcome is 1 bit. If we toss two coins simultaneously, or consecutively, the information is 2 bits, because for the four equal possibilities {tt, th, ht, hh}, we have $H = \log_2 4 = 2$. This additivity holds in general.

For example, when rolling a (fair) *dice* the options are six equal outcomes {1, 2, 3, 4, 5, 6}, so the Hartley information is $H = \log_2 6 \approx 2.58$. However, by rolling a coin and a die at the same time we have 12 pairs {t1, t2, ..., h6} whose Hartley information is $\log_2 2 \cdot 6 = \log_2 2 + \log_2 6 \approx 3.58$. By rolling two dice at the same time it will be $\log_2 6 \cdot 6 = 2 \log_2 6$. The following example discusses the *law of conservation* of Hartley information.

[1]The image does not include cyan, so the wavelengths of the other six have been reprogrammed.

Example 5. *Hartley's information is governed by the conservation law.*

Proof. If two events with N_1 and N_2 equal outcomes occur simultaneously, the information is $\log_2(N_1 \cdot N_2) = \log_2 N_1 + \log_2 N_2$ and is equal to the sum of the information of the individual outcomes. □

However, this additivity for outcomes of different chances is not always sustainable, except for specifically defined "mean values"or in specially selected cases. Their basis is "Shannon information," which is the topic of the sequel.

1.1.3 Shannon's information

For *Shannon information* (1948), the assumption is that given some *probability distribution* p_1, p_2, ..., p_n, which means that each of these numbers is nonnegative and their sum is 1. Then the mean value

$$S = -p_1 \log_b p_1 - p_2 \log_b p_2 - ... - p_n \log_b p_n$$

is called the Shannon information of the given distribution. When these individual probabilities $p_k = 1/N_k$ can be understood as one of N_k equally likely outcomes ($k = 1, 2, ..., n$), then the Shannon information

$$S = p_1 H_1 + p_2 H_2 + ... + p_n H_n$$

is the mean value of the Hartleys $H_k = \log_b N_k = -\log_b p_k$. If the base of the logarithm $b = 2$, then the unit of information is *bit* (BInary digiT), when the base $b = e$ the unit of information is nat (NATural logarithm), and when $b = 10$ the unit is decit (DECImal digiT).

For example, let $p_1 > 0$ be the probability of a "tail" outcome for a (unfair) coin, and let the probability $p_2 = 1 - p_1$ be the probability of a "head" outcome. Similarly, the probabilities of such two outcomes for another coin are $q_1 > 0$ and $q_2 = 1 - q_1$. Then we have the Shannon distributions:

$$S_p = -p_1 \log_b p_1 - p_2 \log_b p_2, \quad S_q = -q_1 \log_b q_1 - q_2 \log_b q_2.$$

These are distributions with $n = 2$ options each. If we toss these two coins simultaneously, or one after the other, there are four outcomes with probabilities $\{p_1q_1, p_1q_2, p_2q_1, p_2q_2\}$ that also form a distribution, because each of these four products is a non-negative number and the sum of all four is unity:

$$p_1q_1 + p_1q_2 + p_2q_1 + p_2q_2 = p_1(q_1 + q_2) + p_2(q_1 + q_2) = p_1 + p_2 = 1.$$

The Shannon information of these four probabilities is

$$S_4 = -p_1q_1 \log_b p_1q_1 - p_1q_2 \log_b p_1q_2 - p_2q_1 \log_b p_2q_1 - p_2q_2 \log_b p_2q2$$

is the sum of the information mentioned in two outcomes. That is the topic of the following example.

Example 6. *From the marked Shannon information it follows* $S_4 = S_p + S_q$.

Proof. Using the previous notations, we calculate:

$$S_4 = -p_1q_1 \log_b p_1q_1 - p_1q_2 \log_b p_1q_2 - p_2q_1 \log_b p_2q_1 - p_2q_2 \log_b p_2q_2 =$$

$$= q_1[-p_1 \log_b p_1 - p_2 \log_b p_2] - p_1q_1 \log_b q_1 - p_2q_1 \log_b q_1$$

$$+q_2[-p_1 \log_b p_1 - p_2 \log_b p_2] - p_1q_2 \log_b q_2 - p_2q_2 \log_b q_2$$

$$= q_1S_p - q_1 \log_b q_1 + q_2S_p - q_2 \log_b q_2 = S_p + S_q.$$

This proves the claim. □

The same applies to tossing an unfair coin with outcomes in the distribution $\{p_1, p_2\}$ and an unfair dice with the distribution $\{q_1, q_2, q_3, q_4, q_5, q_6\}$. Individually, the information is:

$$S_p = -p_1 \log p_1 - p_2 \log p_2, \quad S_q = -q_1 \log q_1 - ... - q_6 \log q_6,$$

and the Shannon information of the joint outcome is:

$$S_{12} = -p_1q_1 \log p_1q_1 - ... - p_2q_6 \log p_2q_6.$$

Choose the base of the logarithm as you wish, so I will not list it.

Example 7. *From the marked Shannon information it follows* $S_{12} = S_p + S_q$.

Proof. Using the previous notations, now for the distribution of 12 options it is:

$$S_{12} = -p_1q_1 \log p_1q_1 - ... - p_1q_6 \log p_1q_6 + p_2q_1 \log p_2q_1 - ... - p_2q_6 \log p_2q_6 =$$

$$= p_1(-q_1 \log q_1 - ... - q_6 \log q_6) - p_1q_1 \log p_1 - ... - p_1q_6 \log p_1 +$$

$$+p_2(-q_1 \log q_1 - ... - q_6 \log q_6) - p_2q_1 \log p_2 - ... - p_2q_6 \log p_2$$

$$= p_1S_q - (q_1 + ... + q_6)p_1 \log p_1 + p_2S_q - (q_1 + ... + q_6)p_2 \log p_2$$

$$= (p_1 + p_2)S_q - p_1 \log p_1 - p_2 \log p_2 = S_q + S_p,$$

and that is the equality sought. □

This idea of "multiplication" of distributions is easily extended to general cases, p with m and q with n options, with the same result $S_{mn} = S_p + S_q$. This is one of the announced "selected" cases when we can say that the conservation law applies to Shannon information. I have investigated other similar ones, and I have listed some interesting examples in the book "Physical Information" (2019). Here is one example from there.

1.1.4 Physical information

In Theorem 2.3.6. of the book Physical Information [11], on page 48, after the binomial, or *Bernoulli distribution* $\mathcal{B}(n,p)$ probability

$$\Pr(B,n,k)=\binom{n}{k}p^k q^{n-k}, \quad k=0,1,\dots,n,$$

where $q=1-p$ is set, the Shannon information is

$$S_n=-\sum_{k=0}^{n}\Pr(B,n,k)\log\Pr(B,n,k).$$

For example, for $n=1$, when we have only one draw, it reduces to

$$S_1=-p\log p-q\log q.$$

Starting from this, $L_1=S_1$, we define the "mean value" called *physical information* for the outcome of n repetitions

$$L_n=-\sum_{k=0}^{n}\binom{n}{k}p^{n-k}q^k\log p^{n-k}q^k.$$

In it, compared to Shannon, the logarithm does not have a binomial coefficient.

Example 8. *The physical information $\mathcal{B}(n,p)$ is $L_n=nL_1$, where L_1 is the information of the simple binomial distribution $\mathcal{B}(1,p)$.*

Proof. Using the previous notations, we calculate:

$$L_n=-\sum_{k=0}^{n}\binom{n}{k}p^{n-k}q^k\log_2 p^{n-k}q^k=$$

$$=-\sum_{k=0}^{n}\binom{n}{k}(n-k)p^{n-k}q^k\log_2 p-\sum_{k=0}^{n}\binom{n}{k}p^{n-k}kq^k\log_2 q$$

$$=-p\left[\frac{\partial}{\partial p}\sum_{k=0}^{n}\binom{n}{k}p^{n-k}q^k\right]\log_2 p-q\left[\frac{\partial}{\partial q}\sum_{k=0}^{n}\binom{n}{k}p^{n-k}q^k\right]\log_2 q$$

$$=-p[\frac{\partial}{\partial p}(p+q)^n]\log_2 p-q[\frac{\partial}{\partial q}(p+q)^n]\log_2 q$$

$$=-np(p+q)^{n-1}\log_2 p-nq(p+q)^{n-1}\log_2 q$$

$$=n(-p\log_2 p-q\log_2 q),$$

i.e. L_n is n times larger than L_1, which was supposed to be proven. □

1.1.5 Boundary information

Having seen that the Shannon information, in some cases, is subject to the conservation law and that it is possible to imitate it with similar physical information for which this law is almost always valid, let us see when the sums of the Shannon information reduce to the probabilities of a given distribution. Otherwise,

$$S = -\int_V \rho(\omega) \cdot \log \rho(\omega)\, d\omega$$

The Shannon information *density of the distribution* $\rho(\omega)$ in the space $\omega \in V$. As we know, these densities are nonnegative numbers, $\rho(\omega) \ge 0$, and collectively $p(\omega) = \rho(\omega) d\omega$

$$\int_V \rho(\omega)\, d\omega = 1,$$

they are unit integral probabilities.

Example 9. *When the discrete probabilities are in a decreasing (non-decreasing) sequence $p_1 \ge p_2 \ge p_3 \ge \dots \ge 0$, or we have a decreasing probability density $\rho(\omega)$ with probabilities $p < e^{-1}$, then the decreasing Shannon information sums are:*

$$-\rho(\omega) \cdot \log \rho(\omega) \ge -\rho(\omega + \varepsilon) \cdot \log \rho(\omega + \varepsilon)$$

for each $\varepsilon \ge 0$.

Proof. Consider the following function $f(x)$ and its derivative $f'(x)$:

$$f(x) = \frac{\log_b x}{x}, \quad f'(x) = \frac{1 - \ln x}{x^2 \ln b}.$$

Since $f'(x) < 0$ for every $x > e$, the function $f(x)$ is decreasing. Substituting $p = 1/x$ we find that $-p \cdot \log_b p$ is decreasing, which was to be proven. □

The conclusion is that smaller probabilities $p \in [0, 1]$, although with larger Hartley information $H = -\log_b p > 0$, make smaller additions and sums to Shannon information. I have given such examples in the blog Surprises and elsewhere. To avoid repeating myself, take a look directly at Exponential II, where there is a comparative graph of the exponential distribution and probability density, respectively:

$$\rho(x) = e^{-x}, \quad \varphi(x) = -\rho(x) \cdot \ln \rho(x)$$

where their closeness is seen otherwise and especially asymptotic approximation, i.e., $\varphi(x) \to \rho(x)$ for $x \to \infty$. We can see that this is true in general in the following example.

Example 10. *For variables $p > 0$, in the case $p \to 0$ the order magnitude of p is $-p \ln p$.*

Proof. It follows easily by applying L'Hôpital's rule (or similar) to the limit:

$$-p\ln p = \frac{\ln\frac{1}{p}}{\frac{1}{p}} = \frac{\ln x}{x} \to \frac{(\ln x)'}{(x)'} = \frac{1}{x} = p, \quad p \to 0.$$

Therefore $-p\ln p \to p$ when $p \to 0$. □

In other words, in situations with small values of the probability distribution, the sum of Shannon information can be considered as probabilities, viewed without the logarithm factor. This introduces the idea of "information of perception"[7], which I will explain in detail later.

1.2 Uncertainty

Information resolves *uncertainty*. In the previous section, we saw that information can be measured, and moreover, that the conservation law may (or may not) apply to this measure. We saw the basic ways of measuring it, and thus we established some quantities of resolved uncertainty. Therefore, we also have measures of uncertainty because the information about the outcome of a random event expresses the amount of prior uncertainty.

1.2.1 Action

A physical action is the product of the change in momentum and path, that is $S = \Delta E \cdot \Delta p$, or the change in energy and time, $S = \Delta E \cdot \Delta t$, expressed in Joule-seconds (J s).

Classically, action (S) is a scalar quantity that describes the change in kinetic (E_k) relative to potential (E_p) energy of a physical system along a path (x) during time $t \in (t_1, t_2)$:

$$S = \int_{t_1}^{t_2} (E_k - E_p)\,dt = \int_{t_1}^{t_2} (\frac{1}{2}mv^2 - mgx)\,dt,$$

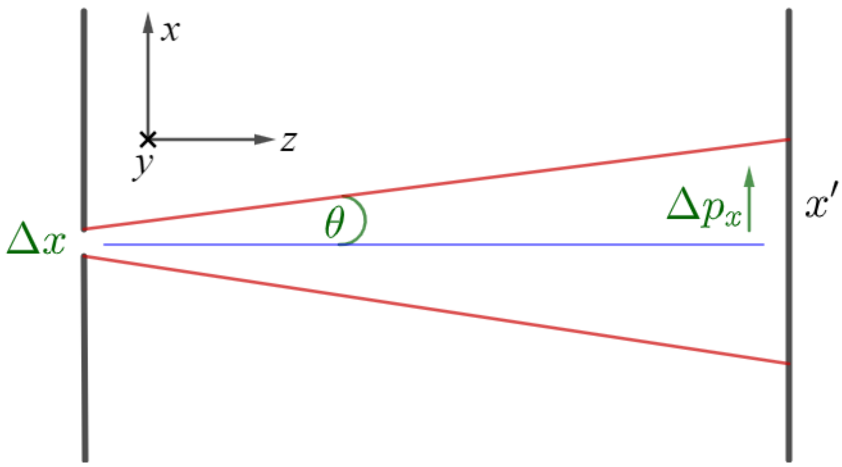
where velocity, $v = v(t)$ and path, $x = x(t)$, are functions of time t. However, when we get down to the very bottom of microphysics, these quantities are *quanta* of the order of Planck's constant. In information theory, quanta are carriers of the smallest portions of physical information, and I consider them equivalent. Simply put, for (this) information theory, a quantum is the limit of certainty or the smallest amount of uncertainty.

1.2.2 Uncertainty relations

Planck's constant $h = 6.62607015 \times 10^{-34}$ J s, as well as the *reduced* Planck's constant $\hbar = h/2\pi = 1.054571817 \times 10^{-34}$ joule-second are measures of the order of magnitude of the limit of a physically real "quantity of uncertainty.". Since the energy

of light is $E = hf$, where the frequency is $f = 1/\tau$, and τ is the period of one oscillation, it is $E\tau = h$. In other words, the product of that small energy and the time of its change is constant. The shorter the time period, the more precisely the moment of the photon is determined, but the greater the uncertainty of the phase in which its energy is located.

Example 11. *When an electron is hit by a photon of shorter wavelength λ, its position $\Delta x \approx \lambda$ is determined more precisely, but the shorter wavelength carries a larger photon momentum $p = h/\lambda$ which, upon collision, increases the uncertainty of the electron's momentum $\Delta p_x \approx h/\lambda$ even more. Hence $\Delta x \Delta p_x \approx h$, the Heisenberg uncertainty relation for position and momentum, is the smallest value of their product.*



In the figure, we see something like the Heisenberg relation of position and momentum of electrons given in the example. Here the particles go from left to right through a narrow slit of height Δx and scatter with a mean angle θ with respect to the initial axis of motion with an uncertainty of momentum Δp_x. Because of the small x' compared to the distance z between the slit and the curtain, the angle θ is small and $x'/z \approx \sin\theta$, and from the figure we see that

$$p_x = p\sin\theta = \frac{h}{\lambda}\sin\theta.$$

We can assume that the probability distribution of light is a Gaussian distribution with position dispersion σ_x and density $\exp(-x^2/2\sigma_x^2)$. The complex amplitude of the wave is then proportional to the root of the intensity $\exp(-x^2/4\sigma_x^2)$ and, after a short calculation, we find $\sigma_x\sigma_{p_x} \approx h/4\pi$, where σ_{p_x} is the Gaussian momentum dispersion. This corresponds to the previous Heisenberg uncertainty of position and momentum, now of the form $\Delta x \Delta p_x \approx \hbar/2$, with the left-hand side of this approximate equality practically always larger.

These are already three ways of understanding the relation of uncertainty of position and momentum. Next comes the proof method, and then we say *principle of uncertainty*, using quantum operators. Before that, let us recall that *standard deviation* σ is a measure of the dispersion of a set X. It tells us how much, on average,

the elements of the set deviate from the arithmetic mean μ of the set. If $\rho(x)$ is the probability density of finding elements in a given set, then the mean and standard deviation are:

$$\mu = E[X] = \int_{-\infty}^{\infty} x\rho(x)\,dx, \quad \sigma = \sqrt{E[(X-\mu)^2]},$$

where $E[X]$ denotes the mean, or expectation, of the elements of the set X. It is easy to find $\sigma^2 = E[X^2] - (E[X])^2$, an expression that is very practical in calculations.

Linear algebra of quantum mechanics works with self-adjoint, i.e. hermitian operators and with statistics but uses its own notation. Thus, we write that a system is in a quantum mechanical state $|\psi\rangle$ that measures two *ospservables* represented by the Hermitian operators A and B, so:

$$A|a_i\rangle = a_i|a_i\rangle, \quad B|b_j\rangle = b_j|b_j\rangle.$$

We write these vectors in Dirac's bra-ket notation. The measurement of A results in one of the eigenvalues of a_i, and the state of the system is extended to all eigenvalues of the operator A so that the above vectors $(i, j = 1, 2, 3, ...)$ form a complete, orthonormal set, so we have:

$$|\psi\rangle = \sum_i \alpha_i|a_i\rangle, \quad \alpha_i = \langle\psi|a_i\rangle.$$

The dot product of the *covariant* vector $\langle\psi| = |\psi\rangle^\dagger$ with contravariant[2] $|a_i\rangle$ is a scalar α_i. The results of the measurement A give a distribution of values with the mean given, notation:

$$\langle A\rangle = \langle\psi|A|\psi\rangle = \sum_i |\alpha_i|^2 a_i.$$

Consistent with these notations, the square of the standard deviation is:

$$\sigma_A^2 = \langle\psi|A^2|\psi\rangle - \langle\psi|A|\psi\rangle^2.$$

This dispersion defines the uncertainty in the measurement A. When we simultaneously measure two observables A and B in the same state $|\psi\rangle$ there will be:

$$\Delta A = A - \langle A\rangle, \quad \Delta B = B - \langle B\rangle$$

the deviations of A from the mean value $\langle A\rangle = \langle\psi|A|\psi\rangle$, and so on with B. The operators ΔA and ΔB are also Hermitian, and the equality of the commutators is easy to check:

$$[\Delta A, \Delta B] = \Delta A\Delta B - \Delta B\Delta A = AB - BA = [A, B].$$

See first example 38, and then the next problem.

[2] See "2.2.5 Dual Spaces" below.

Task 12. *Let's show that the inequality*

$$\sigma_A\sigma_B \ge \frac{1}{2}|\langle[A,B]\rangle| holds.$$

These are general uncertainty relations.

Solution. Consider $|\phi\rangle = (\Delta A - i\lambda\Delta B)|\psi\rangle$, with imaginary unit $i^2 = -1$ and real parameter λ. It will be $\langle\phi|\phi\rangle \ge 0$ due to the Hermitian operators, so:

$$\langle\psi|(\Delta A + i\lambda\Delta B)(\Delta A - i\lambda\Delta B)|\psi \ge 0,$$

$$\sigma_A^2 + \lambda^2\sigma_B^2 - i\lambda\langle\psi|[\Delta A, \Delta B]|\psi\rangle \ge 0,$$

where

$$\sigma_A^2 = \langle\psi|(\Delta A)^2|\psi\rangle, \quad \sigma_B^2 = \langle\psi|(\Delta B)^2|\psi\rangle.$$

Note that $C = -i[A,B]$ is a Hermitian operator, so $-i\langle[A,B]\rangle$ is a real number. Then, from the positive quadratic trinomial (in parameter λ):

$$\lambda^2\sigma_B^2 + \lambda\langle C\rangle + \sigma_A^2 \ge 0$$

it follows that the discriminant of the corresponding quadratic equation is negative:

$$\langle C\rangle^2 \le 4\sigma_A^2\sigma_B^2,$$

$$\sigma_A^2\sigma_B^2 \ge \langle\frac{1}{2i}[A,B]\rangle^2,$$

and hence the required inequality. □

Choosing the position operators $A = x$ and the momentum operators $B = p_x$, we find

$$\sigma_x\sigma_p \ge \frac{\hbar}{2},$$

which is the standard form of the Heisenberg uncertainty relations. The second one relates to time and energy, $\sigma_t\sigma_E \ge \hbar/2$.

Quantum mechanics has become one of the most successful areas of physics and science in general. No other has such a high degree of agreement between theory and measurement, nor such a high precision of predictions, considering that it is actually as much a statistical theory as an algebraic theory. We have no reason to consider its quanta, atoms, and molecules as something non-existent in macro physics, and so let us consider quantum states, vectors, as the reality of the world of larger magnitudes. With the same principle of vector addition and complexity multiplication. And the addition to this is that we will also find general uncertainty relations in the macro world.

Example 13. *Digital media is limited by the amount of recording and by displaying motion; it will be able to have better image resolution (ΔA) with a lower film speed (ΔB) and vice versa. Higher video speed will come at the expense of lower sharpness of details.*

Similar to this example, we notice that as we drive faster, more of the details of the road escape our attention. The life of ants under a tree in the forest we are passing through is invisible to us when viewed from the train compartment, which would not necessarily be so if we were to step out onto the embankment and stop to observe them.

A society with a stronger tradition has fewer other options. We know that the past defines the present and directs the future, from the past positions of particle trajectories that define their future positions, through the evolution of a species that will highlight its limitations, to the biography of a person on the basis of which we expect performance at work.

This is analogous to the development of theories about the future of the universe by observing its past. From example (12) and the above, the saturation of the present (ΔA) and the length of its past (ΔB) are in an indefinite relation: how many times less (more) of one is, so many times more (less) of the other is, if the total information ($[A, B]$) is constant. Hence the importance of these commutators.

1.2.3 Commutators

The idea of a *commutator* can come from the uncertainty relations (Task 12):

$$|\Delta A||\Delta B| \ge \frac{1}{2}|E[[A, B]]|,$$

where $E[X]$ is the mathematical expectation or the mean value of the variable X, and this $X = [A, B] = AB - BA$ is the commutator. When operators are commutative, their uncertainties can be reduced indefinitely (to zero), which indicates the absence of interaction (communication) between the processes A and B, i.e., between the states they represent. The case $[A, B] = \text{const}$ is particularly interesting, which indicates the *law of conservation* during the communication of A and B.

The properties of commutators can be investigated using analytical geometry and the surface they represent. As can be seen from the given link, then $A(x_A, y_B)$ and $B(x_B, y_B)$ are points of the rectangular Cartesian system (Oxy) given by their coordinates. The commutator $[A, B] = x_A y_B - x_B y_A$ is an oriented *surface* of a parallelogram crossed by vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$, the sign of the direction of the circumambulation.

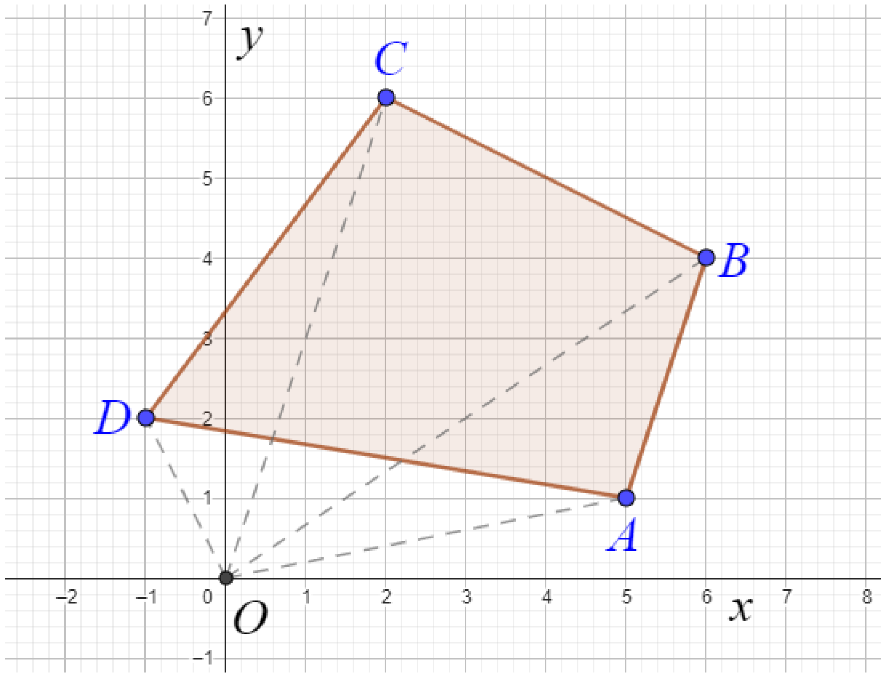
Example 14. *The area of the quadrilateral $ABCD$ in the figure is*

$$[A, B] + [B, C] + [C, D] + [D, A] = 41.$$

Proof. From the figure we read the coordinates of the points: $A(5, 1)$, $B(6, 4)$, $C(2, 6)$ and $D(-1, 2)$. Also, we see that the areas ($ODA = -OAD$):

$$ABCD = OAB + OBC + OCD + ODA =$$

$$= [A, B] + [B, C] + [C, D] + [D, A]$$



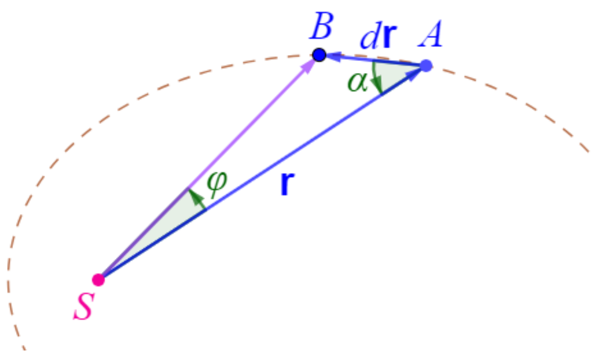
$$= (5 \cdot 4 - 6 \cdot 1) + (6 \cdot 6 - 2 \cdot 4) + (2 \cdot 2 + 1 \cdot 6) + (-1 \cdot 1 - 2 \cdot 5)$$

$$= 14 + 28 + 10 - 11 = 41,$$

and this is the result. □

Example 15. *Commutators can be used in Kepler's second law, so that the motion from the sun to a planet, under the influence of the sun's gravitational field, sweeps out equal areas in equal times.*

Proof. It has been cited several times (Multiplicities, p. 59). I will retell it slowly with the following picture of a constant central force S moving a charge $A \to B$. □



In the figure, a charge travels along a dashed line (conic) from point A to point B driven by a force from point S. The angle $\varphi = \angle ASB$ is so small that the distance AB

is infinitesimal, so the vector $d\mathbf{r} = \overrightarrow{AB}$ is infinitesimal. The point A is determined by the vector $\mathbf{r} = \overrightarrow{SA}$, and the angle α is between $\mathbf{r}$ and $d\mathbf{r}$. Therefore:

$$d\Pi = \frac{1}{2}\mathbf{r} \times d\mathbf{r}, \quad |d\Pi| = \frac{1}{2}|\mathbf{r}||d\mathbf{r}| \sin\alpha$$

the oriented surface of the triangle SAB is also infinitesimal. The choice of coordinate axis orientation changes the sign of the area so that it no longer matters. We divide this equality by the time differential, dt, to get:

$$\dot{\Pi} = \frac{1}{2}\mathbf{r} \times \dot{\mathbf{r}} = \frac{1}{2}\mathbf{r} \times \mathbf{v} = \mathbf{L}/2m.$$

The velocity vector $\mathbf{v} = \dot{\mathbf{r}} = \frac{d}{dt}\mathbf{r}$ is the derivative of the position vector with respect to time. $\mathbf{L}$ is the vector of angular momentum, and m is the mass of the charge, or body A, gravitating around the sun S. After another derivative with respect to time:

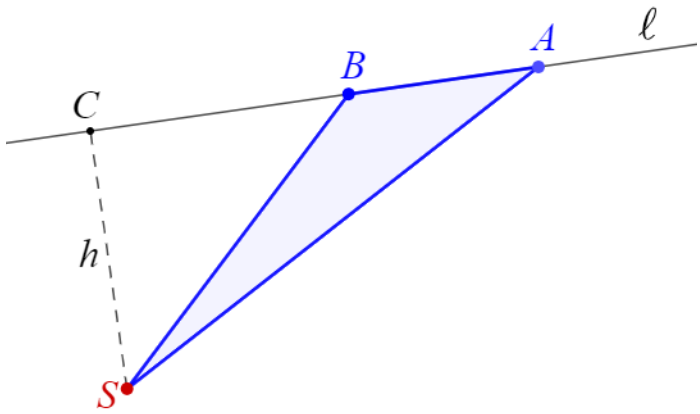
$$\ddot{\Pi} = \frac{1}{2}(\dot{\mathbf{r}} \times \dot{\mathbf{r}} + \mathbf{r} \times \ddot{\mathbf{r}})$$

we find that the first vector product is zero, but so is the second, because the factors are collinear. The first vectors are equal, and the second vectors are in the same direction: position and force. Therefore, $\ddot{\Pi} = 0$, the second derivative of the area of the triangle with respect to time is zero. Hence the second derivative with respect to time is constant $\dot{\Pi} = \text{const}$. This means that the displacement SA, from a constant central force to the charge it moves, sweeps out equal areas in equal times.

The lessons of this calculation are that the (constant) vector of the area $\vec{\Pi}$ swept by the motion SA is proportional to the angular momentum, $\vec{L} = \vec{r} \times \vec{p}$, which means that a body in forced circular motion will spontaneously continue to rotate, the faster with greater momentum ($\vec{p} = m\vec{v}$) the smaller the distance (r) from the constant center of rotation. Second, that this constant area is a real phenomenon, for which the *law of conservation* applies and which we can treat as a physical reality.

Therefore, in a theory that starts from the assumption that information constitutes the weave of space, time, and matter, and that uncertainty is its essence, commutators should also be treated as types of information. The law of conservation of constant surfaces, and thus commutators, becomes equivalent to the law of conservation of information, or uncertainty. Taking action is also equivalent to physical information, from $\Delta E \Delta t = \text{const}$, in the case of constant time intervals, $\Delta t = \text{const}$, the law of conservation of energy will apply and vice versa; from the conservation of energy, the conservation of action and information will follow.

I have often pointed out the reality of these constant surfaces and commutators (Notes I, p. 14), as well as the “spontaneity in forced motion” mentioned here. Discoveries often distort understandings of known phenomena or give them new names. Thus, there is “spontaneity” in the inertial motion of the planets around the Sun, as well as in the increasingly rapid rotation of a ballerina to her tighter hand, but it should also be seen in movements due to the Coulomb force.



In this figure, there is no force at an arbitrary point S. The line ℓ is arbitrary, and along $AB \subset \ell$ it is of constant length, $|AB| = \text{const}$, and it moves along it. However, the area of the triangle $\Pi(SAB) = \frac{1}{2}h|AB|$. This is another consequence of the generalized Kepler's second law, just explained. I mention it to clarify the concept of "spontaneity," which applies equally to strong permanent central forces, whether attractive or repulsive, as well as to their zero values.

Now we learn that bodies continue to move uniformly in a straight line ($A \to B$), by *inertia*, until they are acted upon by other bodies or forces – because they have constant information for the perceptions of surrounding subjects (different ones) that does not change spontaneously. This means that by changing any subject, depending on the power of that subject, a given body will change its previous motion.

1.2.4 Conic sections

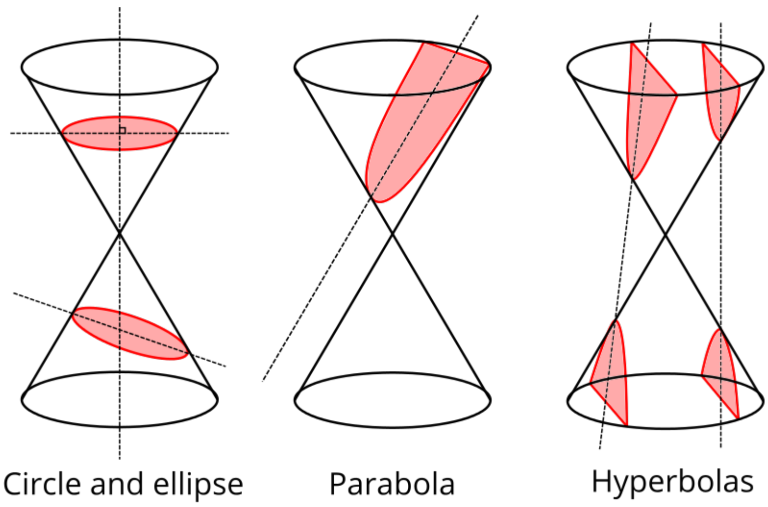
Conics are the intersections of a cone and a plane. In the figure, we see that these are hyperbolas, parabolas, and ellipses with circles as a special case of an ellipse, which also includes a point. Also, special cases are a point and a straight line of contact between the envelope of a cone and a plane. I have been studying conic sections [5] for the sake of teaching mathematics, but also (my) information theory.

In the plane, a conic can be defined by a positive real number ε called the *eccentricity*, a point F called the *focus*, and a line d called the *directrix*. The locus of points T for which the ratio of the distances to the focus and the directrix is equal to the the eccentricity

$$TF : Td = \varepsilon$$

is called a conic. Moreover, if $\varepsilon = 1$ the conic is a *parabola*, if $\varepsilon < 1$ the conic is an *ellipse*, if $\varepsilon > 1$ the conic is a *hyperbola*. This is proven using Dandelin's spheres (Germinal Pierre Dandelin, 1794-1847), in detail in the cited book.

In general mathematics classes, one usually does not go further in explaining conics than their *canonical forms* in the Cartesian rectangular coordinate system



(Oxy), respectively, ellipse, parabola, and hyperbola:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1,\quad y^2=2px,\quad \frac{x^2}{a^2}-\frac{y^2}{b^2}=1.$$

Here $b^2 = \pm a^2(1-\varepsilon^2)$, the plus sign for the ellipse and the minus sign for the hyperbola; for both, the focus $F(c,0)$ is determined by $c = a\varepsilon$, and the directrix $x = d$ is the number $d = a/\varepsilon$. The focus of the parabola is $F(p/2,0)$, and the directrix is $x = -p/2$. In contrast to simple canonical forms, I specifically mention the general second-order curve and the usually avoided proof.

Theorem 16. *The equation of a conic in the Cartesian system is a second-order curve:*

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0.$$

Proof. Let $\varepsilon > 0$ be the eccentricity, the point $F(c_x, c_y)$ the focus, and the line

$$d : mx + ny + p = 0$$

the directrix of the conic. Its arbitrary point is $T(x,y)$. Hence:

$$TF : Td = \varepsilon,$$

$$\sqrt{(x-c_x)^2+(y-c_y)^2} : \frac{|mx+ny+p|}{\sqrt{m^2+n^2}} = \varepsilon.$$

The well-known formula for the distance of a point from a point and a point from a line is used. Next, we square this and arrange:

$$(x-c_x)^2+(y-c_y)^2 = \varepsilon^2\frac{(mx+ny+p)^2}{m^2+n^2}.$$

Comparing with the given expression, we find:

$$a_{11} = m^2(1-\varepsilon^2)+n^2,\quad a_{12} = -2mn\varepsilon^2,\quad a_{22} = m^2+n^2(1-\varepsilon^2),$$

$$a_{13} = -2c_x(m^2+n^2) - 2mp\varepsilon^2,\quad a_{23} = -2c_y(m^2+n^2) - 2np\varepsilon^2,$$

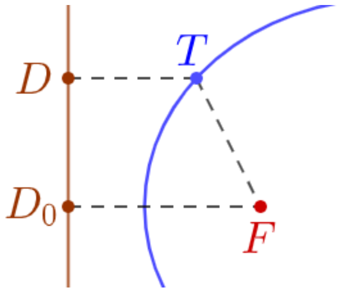
$$a_{33} = (c_x^2+c_y^2)(m^2+n^2) - p^2\varepsilon^2.$$

Therefore, every conic is a second-order curve. □

Those second-order curves that are not ellipses, parabolas, or hyperbolas are called degenerate curves, and these can only be lines, points, and empty sets. Degenerate curves are also called *degenerate conics*.

In a polar *coordinate system* , a point $T(r,\phi)$ in a plane is determined by the distance $r = OT$ from a fixed point O of that plane, which we call the pole, and the oriented angle $\phi = \angle(xOT)$ of the point with respect to a fixed axis, the x-axis of that plane. Let $T'(x,y)$ be the same point in Cartesian coordinates, then:

$$\begin{cases} x = r\cos\phi, & y = r\sin\phi, \\ r = \sqrt{x^2+y^2}, & \phi = \operatorname{arctg}\frac{y}{x}. \end{cases}$$



In the figure, $T(r,\phi)$ is a point of the conic; its focus $F = O$ is also the pole, the origin of the polar coordinate system, and the line D_0D is the directrix.

Theorem 17. *When the focus is at the pole, and the directrix is perpendicular to the polar axis, in the image D_0F, at a distance of p from the pole, then the equation of the conic in the polar coordinate system is*

$$r = \frac{\varepsilon p}{1-\varepsilon\cos\phi},$$

where $\varepsilon > 0$ is the eccentricity of the conic, $\varepsilon < 1$, $\varepsilon = 1$ and $\varepsilon > 1$ for the ellipse, parabola, and hyperbola, respectively.

Proof. In the previous figure, the focus of the conic is at the pole, which is at a distance p from the directrix DD_0 perpendicular to the polar axis FD_0 at the point $D_0(p,\pi)$. The point $T(r,\phi)$ belongs to the conic, and the point D of the directrix closest to it is therefore:

$$TF : TD = \varepsilon, \quad TD = TF/\varepsilon = r/\varepsilon.$$

However, the vector equality also holds:

$$\overrightarrow{DD_0} + \overrightarrow{D_0F} + \overrightarrow{FT} = \overrightarrow{DT}.$$

In the Cartesian system, the abscissas of these four vectors are respectively 0, p, $r\cos\phi$ and r/ε, which applied to this equality gives $p + r\cos\phi = r/\varepsilon$, and hence $r(1-\varepsilon\phi) = \varepsilon p$ and the required equality. □

In the case of an ellipse (+) and a hyperbola (-), by comparing with the canonical forms, we easily find:

$$a = \frac{\varepsilon p}{\pm(1-\varepsilon^2)}, \quad b = \frac{\varepsilon p}{\sqrt{\pm(1-\varepsilon^2)}}, \quad c = \varepsilon a, \quad d = \frac{a}{\varepsilon}.$$

In the case of a parabola, this p is also a parameter of its canonical form.

It is shown that conics are trajectories of motion of charges under the action of permanent central forces. To solve these problems, theoretical physics often uses the *Euler-Lagrange equations* instead of Newton's method:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) = \frac{\partial L}{\partial q},$$

where q gives the line of motion. In the Cartesian system, q is the abscissa (x), the ordinate (y), or the applicate (z), in the polar system it can be the distance (r), or the angle (ϕ). The difference between the kinetic (E_k) and potential energy (E_p) is the *Lagrangian*:

$$L = E_k - E_p.$$

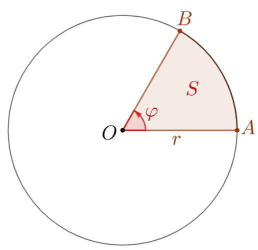
Look at the school examples first (Space-Time, [8], 1.2.6 Lagrangian) to make it easier to understand what follows.

Example 18. *The Euler-Lagrange equations of the central force are:*

$$\frac{d}{dt}(mr^2\dot{\varphi}) = 0, \quad m\ddot{r} - mr\dot{\varphi}^2 = f(r).$$

The mass of the charge is m, the distance from the center is r, the polar angle is φ, and $f(r)$ is the force.

A circular section OAB, or sector, as in the figure on the right, has an area $\frac{1}{2}r^2\varphi$ and a sector velocity $S = \frac{1}{2}r^2\dot{\varphi}$. In the case of motion ($A \to B$) under the action of a central force from point O, as we saw when calculating Kepler's second law, this velocity is constant, so



$$\frac{d}{dt}(mr^2\dot{\varphi}) = 0, \tag{1.1}$$

where m is the mass of the rotating body. Even when it is not a circle, a charge (planet) moves under the action of a central force (around the sun) so that in equal times its radius vector ($OA \to OB$) sweeps out equal areas. This means that the sector velocity does not change.

On the other hand, the kinetic and potential energies of such a constraint are:

$$E_k = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\varphi}^2), \quad E_p = -\int f(r)\,dr,$$

so the Lagrangian is

$$L = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\varphi}^2) + \int f(r)\,dr.$$

Putting this into the Euler-Lagrange equations, replaced with $q = \varphi$ we find the previous (1.1) constant sector velocities. Substituting $q = r$ we find

$$m\ddot{r} - mr\dot{\varphi}^2 = f(r). \tag{1.2}$$

These are (1.1) and (1.2), the Euler-Lagrange equations of the central force of the example.

The first of the equations (1.1) is, I hope, explained. The second, (1.2), needs a little more explanation. This will be helped by eliminating time from the two in order to find the "equation of the trajectory.". Let us note that this is a general problem of motion under the action of a central force and therefore very interesting to us.

We start from the equations already obtained:

$$\ddot{r} - r\dot{\varphi}^2 = \frac{1}{m}f(r), \quad r^2\dot{\varphi} = 2S,$$

so by substituting the second into the first, we find:

$$\ddot{r} - \frac{4S^2}{r^2} = \frac{1}{m}f(r). \tag{1.3}$$

By integrating this equation, we would obtain the final equation of the form $r(t)$. That's already something, but let's look for more and apply the transformation:

$$\ddot{r} = \frac{d\dot{r}}{dt} = \frac{d}{d\varphi}\left(\frac{dr}{d\varphi}\frac{d\varphi}{dt}\right)\frac{d\varphi}{dt} = \frac{d}{d\varphi}\left(\frac{2S}{r^2}\frac{dr}{d\varphi}\right)\frac{2S}{r^2} = -\frac{4S^2}{r^2}\frac{d^2}{d\varphi^2}\left(\frac{1}{r}\right).$$

This transformation and the previous equation give:

$$-\frac{4S^2}{r^2}\frac{d^2}{d\varphi^2}\left(\frac{1}{r}\right) - \frac{4S^2}{r^2} = \frac{f(r)}{m},$$

$$\frac{d^2}{d\varphi^2}\left(\frac{1}{r}\right) + \frac{1}{r} = -\frac{r^2 f(r)}{4mS^2}. \tag{1.4}$$

This is the differential *equation of the trajectory* (Binet's formula). The equation is very general for central forces from the pole and, at the same time, very practical for further studies of motion under force from that center.

Task 19. *Let us show that the solution of the general trajectory equation (1.4) is a conic.*

Proof. We start from (Theorem 17) the equation of the conic in the polar system:

$$\frac{1}{r} = \frac{1 - \varepsilon\cos\varphi}{\varepsilon p},$$

$$\frac{d}{d\varphi}\frac{1}{r} = \frac{\sin\varphi}{p},$$

$$\frac{d^2}{d\varphi^2}\frac{1}{r} = \frac{d}{d\varphi}\left(\frac{\sin\varphi}{p}\right) = \frac{\cos\varphi}{p},$$

then using the conic equation and (1.4), we find:

$$\left(-\frac{1}{r} + \frac{1}{\varepsilon p}\right) + \frac{1}{r} = -\frac{r^2 f(r)}{4mS^2},$$

$$f(r) = -\frac{4mS^2}{\varepsilon p}\frac{1}{r^2}.$$

This is the central force that decreases with the square of the distance and moves charges along conic paths. □

Thus, permanent central forces decrease with the square of the distance, move charges along the trajectories of ellipses, parabolas, or hyperbolas, and the pull to them sweeps out equal areas at equal times. Gravity is an example of such, Coulomb's law is another example, but there are third ones. Permanent here means a conservation law that gives these forces a kind of reality and, to us, the possibility of their deeper connection with information.

1.2.5 Probability force

The topic of this introductory story is one of my theoretical research, for which I also did some statistical tests (Uncertainty Force, 2022). The basic idea is to see the more frequent occurrence of more probable outcomes in the realizations of random events as the action of some "force.". To compare such with physical forces and find their similarities or differences. The experiment is, of course, completely outside the current knowledge of science, but it is mature enough (in my works) and interesting enough to mention.

The problem with probability is that it is much more exact than it seems. It is a slippery slope for the superficial, the hasty, or the careless, and experts often get into arguments with laymen in which they end up "losing." Here is an example from my practice, a supposed argument over the Internet.

On a social network, I inadvertently "liked" and posted on my "wall" a problem that at that moment seemed like a nice example for an introduction to "conditional probabilities," a teaching topic that awaited me.

Example 20. *A father has two children. It is known that one is male. What is the probability that the other is also male?*

Solution. We assume a 50/50 chance that the child will be male (man) or female (woman). Then the following four options are equally likely: {mm, mw, wm, ww}. In only three cases of these pairs is it male, and in only one of those three is the other male. Therefore, the required probability is 1/3. □

I did not write the solution when I posted the task, believing it to be trivial, but the next day I found myself in long and mostly naive discussions. Many people thought it was "logical" that each subsequent child always had a 50/50 chance of being male or female, so they thought the answer was 1/2. Those "perfectionists" were concerned with the exact probability of birth, and this number varied from 0.4 to 0.6, depending on the depth of knowledge of birth rates in certain regions of the world.

When I published the solution, there were a few colleagues from the same field on my side, but we were met with an odium of corrections, attacks, and even bad words. It's a shame that you morons are teaching mathematics to our children! – It was at my (our) expense. This country is failing with people like that, phew! – They got angry.

I don't usually interact with idiots, and due to the exclusion of some from my network, I don't know the state of this discussion below, but its beginning heralds a point that is often repeated to us and which is sometimes used for good. People are naive about probability, and tests of mathematical expectation (mean value), dispersion (average deviation from the mean value), and deviation given the length of a series of random outcomes can detect many attempts at deception.

Example 21. *A coin is supposedly tossed 24 times and the outcome is the sequence (t, t, h, h, t, h, h, t, h, t, t, h, t, h, t, t, t, h, t, t, h, t, t, t). How random is this sequence?*

Solution. A fair coin has a probability $p = 1/2$ of falling "tails" and the same number $q = 1 - p$ of falling "heads.". In a given sequence of $n = 24$ tosses, "tails" fell $k = 15$ times, and "heads" $n - k = 9$ times. There are such distributions:

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} = \frac{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6}{(1\cdot 2\cdot 3\cdot 4)(1\cdot 2)} = 15$$

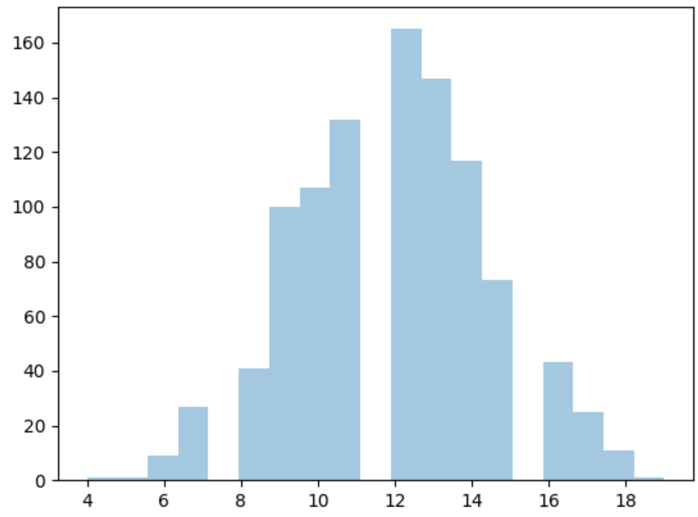
and the probability of each is $p^k \cdot q^{n-k}$. A sequence of $n = 24$ tosses without memory is a Bernoulli trial $\mathcal{B}(n,p)$, part of the binomial distribution probability:

$$\Pr(B,n,k) = \frac{n!}{(n-k)!k!} p^k q^{n-k}, \quad k = 0, 1, ..., n$$

that in n coin tosses, exactly k heads will appear. The expected value μ and the variance σ^2, i.e., the mean square deviation from the expected value, are:

$$\mu = E[k] = np = 12, \quad \sigma^2 = E[(k-\mu)^2] = npq = 6.$$

We can't learn much from just one such short sequence, but look at the following graph for a simulation of 1000 such tosses. On the lower axis (abscissa) are the numbers $k \in \{0, 1, 2, ..., 23, 24\}$, and on the vertical axis (ordinate) is the number of occurrences of that $k = 15$ out of $n = 24$ in 1000 series. About a third of all outcomes are those k from $\mu - \sigma = 9.55$ to $\mu + \sigma = 14.45$, because the dispersion $\sigma = \sqrt{6} \approx 2.45$. Therefore, the given series could be random. □

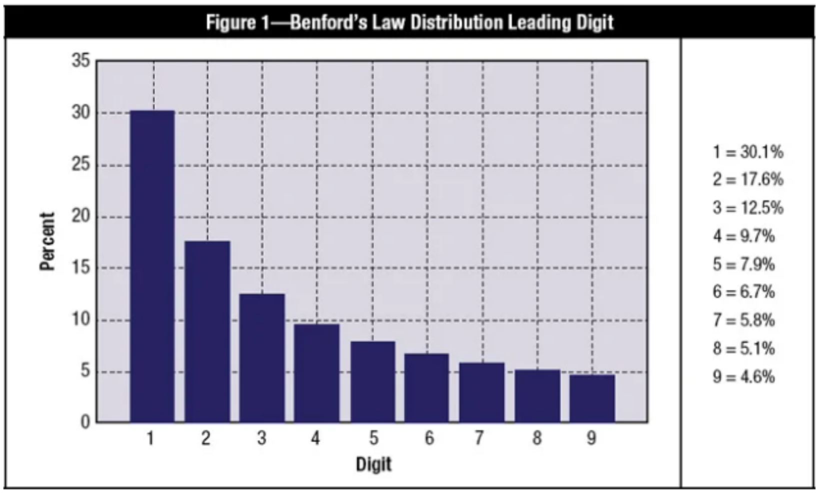


If a person were to say “tail” or “head” by heart in an attempt to imitate randomness, they would probably miss this limit, which, as we can see, changes with changing n. Deviation from this limit, which for an increasing number of simulations looks more like the otherwise complicated Gaussian “bell” of the normal distribution, indicates fraud. For example, it is possible to test lotteries around the world, which they often prevent by not publishing either the number of hits or the number of tickets paid for (which is again suspicious in its own way).

One interesting case happened to me in a company when they needed to check the work of the accounting department. "Give me the keys to the cabinets and archives,"I was told, and rummage through tons of folders to your heart’s content. And we won’t do that—I told them, confused—but give us the numbers themselves, any numbers from the accounting records, but especially by department. Do you want the numbers themselves, without any accompanying explanations, so that we can just pile them up and sort them by quarter? – The employees looked at each other in astonishment and commented that someone must be crazy here. A person who does programming and knows anything about economics or accounting wouldn’t be blabbering like this—I heard quiet gossip.

However, I used the less well-known Benford’s Law (Frank Benford, 1938) of the occurrence of leading digits of numbers. It predicts a distribution of the first digits, as in the following figure (Leading Digit), which is very difficult to achieve by writing numbers by heart or by falsifying numerical data for the desired purposes. It is difficult to describe all the confusion of the employees when the method singled out a section with an error, which was confirmed and yet turned out to be unintentional during the entry of innocent accountants.

These tests can meanwhile detect falsifications of data from various numerical publications and, for example, leave those who commissioned the check convinced that you have insiders there. So great is the ignorance of "experts"in probability about its processes that even your admission about the way it works will not dissuade them.



Let us now return to the announced interesting facts of probability theory for an introduction to (my) information theory. First, let us note that deviations from expectations are also random events, again with computable values. When we work with random numbers, usually labels x from the set X, such as the number k of "letters" from the set {0, 1, ..., 24} in the example, then it is useful to know two expectations:

$$E[X] = \sum_{x\in X} x \cdot \Pr(x), \quad E[X^2] = \sum_{x\in X} x^2 \Pr(x).$$

In this case, random numbers (x) can be negative and discrete, and there can be a continuum when we use integrals. In any case, the following *Chebyshev's inequality* (1867) holds.

Theorem 22. *A random variable X with absolute values $|x_1|, ..., |x_m|$ in some interval of positive real numbers (a, b) and finite mathematical expectation of the square $E[X^2]$, for each $r \in (a, b)$ satisfies Chebyshev's inequality for probabilities:*

$$\Pr(|X| \ge r) \le \frac{E(X^2)}{r^2}.$$

Proof. Let $p_k = \Pr(x_k)$, so we have:

$$E(X^2) = \sum_{k=1}^{m} x_k^2 p_k = \sum_{x_k<r} x_k^2 p_k + \sum_{x_k\ge r} x_k^2 p_k \ge \sum_{x_k\ge r} x_k^2 p_k \ge \sum_{x_k\ge r} r^2 p_k = r^2 \Pr(X| \ge r).$$

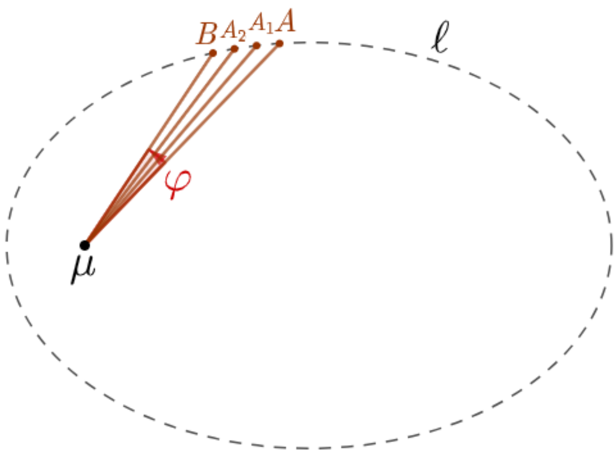
Let's divide this by r^2 and there is the required inequality. □

Here is an example of using this inequality in practice. Suppose that an asset is randomly selected from a population of assets. The population can be of any statistical sample, and we only know that its average return is $\mu = 12\%$ and its variance is $\sigma = 3\%$. To find the probability that an asset selected at random from this population has a return less than 8% or greater than 16%, we can apply Chebyshev's inequality:

$$|X-\mu| \ge r, \quad r = 4\%,$$
$$\Pr(|X-\mu| \ge r) \le \frac{\sigma^2}{r^2} = \frac{3^2}{4^2}$$
$$\Pr(|X-\mu| \ge r) \le \frac{\sigma^2}{r^2} = 56,25\%$$

Therefore, the probability that an asset's return will be less than 8% or greater than 16% of the population of assets, which has a mean return of 12% with a variance of 3%, is less than 56.25%. We find this from Chebyshev's inequality, with very limited representations of the sample, knowing only three numbers μ, σ, r.

What interests us here is only the decrease in probability with the square of r when we consider random variables from the set X further away from their expected value μ. The dispersion σ is a peculiarity of the distribution itself, say like the mass M of a planet, or the sun in Newton's gravitational force GM/r^2. Abstract probability spaces thus become similar to spaces of central forces, with other consequences of the same kind.



For example, the motion from the sun to the planet sweeps out equal areas in equal times (Task 19). We can understand this from this figure by the fact that the total force over the elapsed time, along the line ℓ of motion of the points $A, A_1, A_2, ..., B$, is equal. This is achieved by keeping the point longer at places further from the center of force μ, where the force is weaker. The result is equal areas that sweep out the stretch $A \to B$ in equal times, with the points drawing the line ℓ of the conic. The form is the same in the case of gravity, so the result of calculating these integrals is the same. In the given figure, μ is at the focus of the ellipse ℓ.

The abstractness of this space of probability is neither less nor greater than the abstractness of gravity or matter itself. We know that Louis de Broglie (1924) put forward the hypothesis of *matter waves*, which served as the basis for *Schrödinger* (1926) to compose his now widely known equation of quantum mechanics. It has proven to be more than good in predicting experiments, since when, together with

$A \vee B$	$\top$	$\bot$
$\top$	$\top$	$\top$
$\bot$	$\top$	$\bot$

$A \wedge B$	$\bot$	$\top$
$\bot$	$\bot$	$\bot$
$\top$	$\bot$	$\top$

the matrix mechanics of Heisenberg (1925), we consider it the main source of understanding that micro-world.

In the considerations of information theory, I hold that is *real*, that which lasts, more precisely that for which the *law of conservation* holds, some forms of which we encounter in physics. In addition, *truths*, *physical laws* also last, so they are also some forms of reality. With this, information theory becomes broader than physics but remains in the domain of mathematics. However, when investigating untruths, it will seem to us that both of these frameworks are narrow for information theory.

1.3 Truth

This will not be a story about Boole's (1847) algebra, nor about electric circuits, but rather their interesting continuation. But the importance of that part of the algebra of logic deserves to be mentioned in some parts. Let us note that hence the concept of *statement* as a mathematical sentence that can be either true ($\top$) or false ($\bot$), negation, which translates true into false ($\neg\top \to \bot$) and vice versa. Of the binary operations, these are *disjunction* $A \vee B$ which gives a false statement if and only if both statements A and B are false, then the conjunction $A \wedge B$ which gives a true statement if and only if both statements A and B are true.

Every other operation can be reduced to these three. For example, the implication $A \Rightarrow B \equiv \neg A \vee B$, because it is false if and only if the premise (A) is true and the consequence (B) is false. In general, when we have some statement $f(A, B)$ that is true only when A is true and B is false, or conversely when A is false and B is true, we have $f(A, B) = (A \wedge \neg B) \vee (\neg A \wedge B)$. In this way, using the disjunctive form, we can form statements with arbitrary variables. The same is possible with the conjunctive disjunctive form.

1.3.1 Symmetry

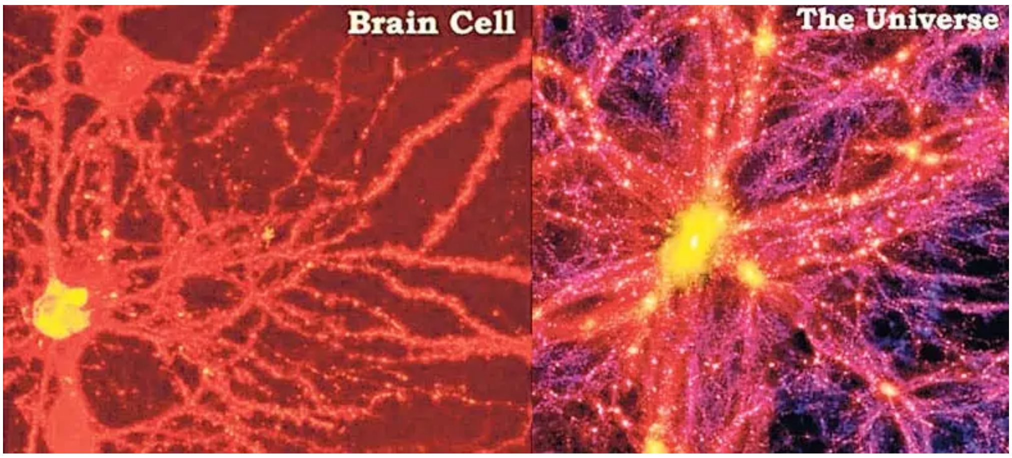
There is a simple and all the more strange because it is an unexplored symmetry of "true" and "false" with respect to negation ($\neg$). By negating correctness, one obtains incorrectness; by negating disjunction, one obtains conjunction; and by negating *tautology* (always a true statement), one obtains *contradiction* (always a false statement). This *bijection* (mutually unambiguous mapping) means the equivalence of šacred truth"and "sacred lies," which will be discussed in more detail below.

For example, the tautology $A \vee \neg A$ becomes the contradiction $\neg A \wedge A$ by negation, and the following disjunction table becomes the conjunction table by negation:

This mapping is symmetric, so the reverse is also true: the negation of a conjunction gives a disjunction, the negation of a contradiction gives a tautology, and so on. Whatever the quantity of truths (determined by some measure) is, the quantity of falsehoods is equal to it. Both are from infinite domains, and hence the following problem.

According to the Löwenheim-Skolem theorem, if a first-order theory (as in mathematics, philosophy, linguistics, and computer science), sentence, or formal system has any model, it has a countable model (whose members can be paired with positive integers). Moreover, it has a model for every cardinal number greater than or equal to the countably infinite $\aleph_0$. Consequently, first-order theories cannot control the cardinality of their infinite models.

With the aforementioned symmetry, this seizure of control is carried over from truths to falsehoods and vice versa, with interesting consequences. *Infinity* is defined by a set that can be a proper subset of itself. For example, the set of natural numbers $\mathbb{N} = \{1, 2, 3, ...\}$ is a proper subset of $\mathbb{N} \subset \mathbb{Z} = \{..., -2, -1, 0, 1, 2, ...\}$, of the set of integers, both of which are of cardinality $\aleph_0$. Moreover, the set of integers is a proper subset of the set of rational numbers, $\mathbb{Z} \subset \mathbb{Q} = \{\frac{m}{n} \mid m \in \mathbb{Z},\ n \in \mathbb{N}\}$, also countably infinite (Sets). However, there are other cardinalities, such as the set of real numbers $\mathbb{R}$. Its continuum is of the label $\mathfrak{c}$ and is greater than $\aleph_0$. The set $\mathbb{R}$ also has real subsets of the same cardinality as, say, the interval $(0, 1) \subset \mathbb{R}$.



One of the consequences of Skolem's theorem is described in the image above, with brain cells on the left and parts of the universe on the right. Science: "Compared images of brain cells and the cosmic web of galaxies make it difficult to distinguish the two. So the universe can appear to us as one giant brain, or vice versa, as if there is a small universe in each of our brains. It's not just some fun thought. In a new study, an astrophysicist and a neurosurgeon have documented striking similarities between the cosmic web of galaxies and the neural networks of brain cells...".

However, we will not follow the fantastic comparisons of the brain and the cosmos of the cited article from a famous scientific journal any further, but rather we see here the consequences of theoretical models in general. Theories, whether

whole or in parts, appear in countless practices. As we will soon see, this is the opposite of unique concrete phenomena. Infinity can always be added to something greater or lesser without it remaining the same, which is not the case with finite things. Perceptions are finite. They share with these a truth to which they then give *reality*.

Colloquially, we say that something is *real* when it is true and sustainable. For example, to say "This is a realistic assessment of the situation," or "The organization of society is realistic," or "These are real possibilities," means that "the assessment is valid," or "the organization is sustainable," or "these possibilities are feasible." It turns out that the concept of physical reality brings this under the "law of conservation," which, as in the case of energy, allows its changes into various forms (potential, thermal, kinetic, chemical, and similar energy) but does not allow its total quantity to change.

We will add here: when something lasts, then it has a past. It is growing with respect to the current present, so we have a *paradox of persistence*. If we do not include the past in the "totality," then we do not have something that "remains the same," and if we include the past, then the present becomes thinner and more diluted due to past events. This is an important point in my theory of information. Past events affect the present, directing and limiting them in such a quantity that the totality of the past and the present is constant.

The said "action" is the sending of information, due to which the past fades and becomes thinner, all the more so the older it is. This does not mean that the present does not age. However, within our perceptions, its age constantly remains within the framework of finality. The aforementioned paradox, therefore, resolves the past that acts unidirectionally on the present, maintaining its totality, while the older one continues to become thinner and longer, reducing the uncertainties of current and future events.

On the other hand, physical nature cannot make something untrue. When we prove that something is not possible, it cannot happen. The consequence is that the "world of lies" is not part of inanimate matter and that it is not permanent. Given the aforementioned symmetry of truths and lies, it would then be incomprehensible for lies to decay and disappear while being equivalent to permanent truths if there were no infinity in their structure. Lies in the domain of our reality are constantly being created, decaying, and disappearing forever, but they maintain some kind of balance with the "quantity" of the present.

This is consistent with the increasing certainty of future events by exhausting the information of the present in order to stretch the past so that there is less and less truth as there are fewer lies. The first say that the world is increasingly ordered, but in such a way that more laws merge into fewer of them, and the second, given the *vitality*, which is the truth added to the lie, that there is less and less room for *living beings*.

1.3.2 Vitality

A dead physical being cannot lie, does not recognize lies, and does not react to them. However, in some situations it is attached to lies when I call this addition to the truth vitality. Due to the just described "equal amount" of truth and lies in the environment, the question of their non-interference in "normal" (read: lifeless) situations is puzzling. There are at least two explanations for this behavior. The first is that not everything communicates with everything, and the second I would describe as the "photoelectric" effect. So much for the beginning.

You cannot read a text if you are illiterate, nor understand speakers of other languages, or simply enter someone's code without a key. It is similar to the sense of sight or hearing, which can only perceive frequencies, light, or sound of a certain range. Everything in nature is reduced to waves, and then this prohibition of communication. Like a ball in a funnel that cannot jump out without sufficient speed, but with a strong momentum can find itself outside the recess, so half-truths cannot be thrown to the level necessary for the information of a dead physical substance.

One explanation is also possible using Parrondo's paradox. We find it in combinations of losing games that give a winning game. Let's say that with the associated state A, a given system S loses 1 point of some "action"with each subsequent step (interval of time). It is irrelevant whether the system S notices this change at all. When it is associated with state B, it reacts to the stock of "action,šo when it is an even number, 3 units of it are added, and if it is odd, 5 units are subtracted. With the state C, it alternately flip-flops, gaining 1, then losing 1.

Continuing with only A or only B, the system loses 100 points after exactly 100 moments (steps). Sticking with only C, after $n = 100$ steps it will remain with the initial value because the probabilities ($p = q = 1/2$) of winning and losing are equal, with variance $\sigma^2 = npq = 25$, i.e., dispersion around the mean value $\sigma = 5$. This is a binomial distribution.

However, if the system alternates between the first and second states in the sequence ABABAB..., and starting with 100 points, after the first step A has 99 points left and in the first step B loses another five. He has 94 points left, again an even number, which is why in the next cycle AB loses another six points, and again an even number, 88, remains. After 16 cycles, AB loses $16 \cdot 6 = 96$ points, so if it continues, it goes into the negative.

Conversely, when it alternates with the second and then the first state, descending BABABA..., after the first B it has 103 points. With the first A, it loses one and is left with 102 points. This is also an even number, which is why in the next cycle BA gains two more points. After each cycle BA gains two points so that after $k =$ 1, 2, 3,... cycles BA will have $K = n_0 + 2k$ points. After k = 50 cycles and an initial $n_0 = 100$ points, it reaches $K = 200$ points. This combination of two losing joints has become a winning one.

It was Parrondo's combination of losing games that resulted in a winning one. It irresistibly reminds us of rhythmic synchronization, of waves, and further of the explanation of non-communication by the photoelectric effect. However, not all

the explanations that I could give here, nor all those that I would know how to give, are presented here. I will mention only one more from the field of linear algebra because of the importance of that part of mathematics in applications.

Let's look at the simulation of getting "something" from "nothing" using the following two matrices that map the to the *zero space* of the algebra:

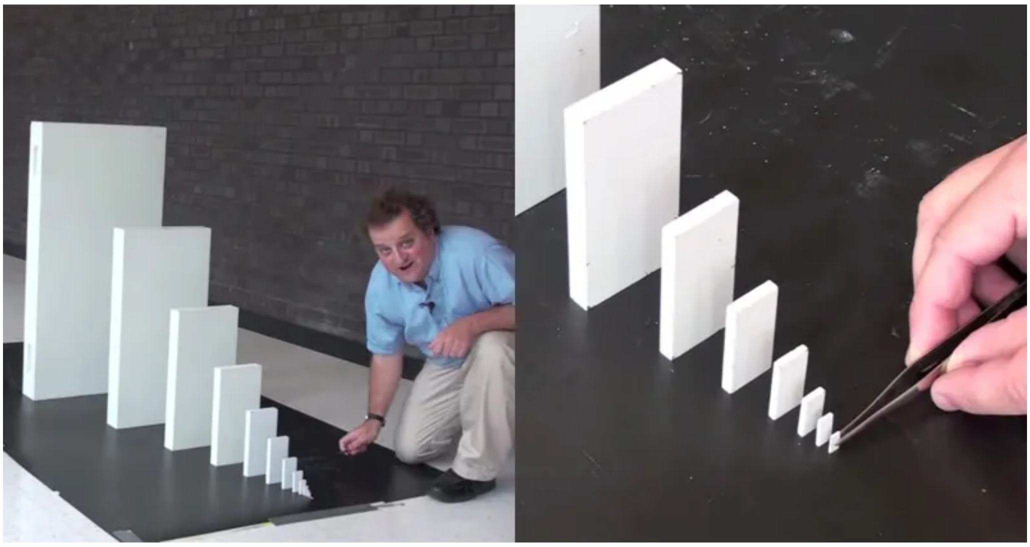
$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

It is easy to check that multiplying each of these matrices by itself gives the zero matrix, $A^2 = 0$ and $B^2 = 0$, which means the non-stationarity of the processes they represent. These matrices represent *phantom processes*, which disappear when isolated. But they last combined, joined together:

$$AB = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad BA = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Thus $(AB)(AB) = (AB)^2 = AB$, a process that does not die out, and so does $(BA)^2 = BA$. Moreover, their sum is the unit operator, $AB + BA = I$, isometry, i.e., an operator that preserves norms, i.e., a process for which the conservation law applies. Here, in translation, it would be an example of obtaining reality from unreality.

One way or another, a lie manages to act on an inanimate object through life. It is this vitality that, by thus attaching itself to permanence, manages to prolong its duration and in return gives substances an excess of options and freedom. Let us note that weak information, in this combination of lies and truth, can produce a chain reaction of great power.



We are familiar with the domino effect in many macro natural phenomena, for example, meteorological, or in the mathematical theory of deterministic chaos. A small movement of the index finger on the trigger of a rifle will cause a terrible

shot and perhaps an accident. It is not surprising that the aforementioned multiplication of theoretical schemes (Skolem's theorem) has its manifestations in living beings.

1.3.3 Incompleteness

Gödel (1931) successfully dealt with questions of truth and falsity in a special way, inspired by set theory and Russell's paradox (1903), which became one of the main topics of mathematics in the study of logic. He socialized with the greats of his time and at one time was personally very popular and influential in wider such circles. The foundations of Gödel's discoveries are the theorems on *incompleteness* which are very difficult in their original form, but I hope that I can retell their point and prepare it for consequences in (my) information theory. Here is what it is about.

A theory is *consistent* if it is so correct that it does not prove falsehoods, that it is based on consistent axioms, and *complete* if its axiom system is sufficient for the completeness of the theory. Gödel's first incompleteness theorem states that a consistent theory cannot be complete, and the second theorem, which is an extension of the first, states that a system cannot establish its consistency. Gödel's incompleteness theorems were the first of several closely related theorems on the limits of formal systems.

Let us try to understand the first of these theorems using a "deductive machine" (Quine, 1946). Let us imagine such a machine to which we can give and from which we can receive truths and only truths. It can give us back, say "two plus two is four," but it cannot say "two plus two is five." We can say that this machine cannot say two plus two is five, but not it, because in doing so it would utter a sentence that is not true. The machine cannot say, "I cannot say two plus two is five," because that sentence contains an untrue statement.

So, a lie is needed to bridge the gap to some kind of "completeness" and reach those truths that we could not reach by adhering to the truth and only the truth. An example is mathematics itself, which reveals pure truths but has in its tools also lies, for example, in the method of proof by *contradiction*. This is, I remind you, the procedure when we start from an assumption and (by good) deduction we derive two opposite, contradictory statements, A and $\neg A$, from which the conclusion follows that the assumption was false.

From the point of view of information theory, this is an important discovery about the difference between mathematics itself and inanimate nature. If inanimate nature could know what we know about it, it would not be what it is. It would have the power to lie, which it does not have. Again, by building truth with lies, we have no guarantee of the exact equality of these two structures, the building of mathematics and pure natural truths. There is also a result of the second incompleteness theorem, that a system cannot see its own consistency: if it is consistent, it is falsifiable, and if it is not, then there is nothing to determine.

Given the lies that are exclusively our contribution, the consequence of the vi-

tality or excess of options that we possess in relation to the substances of which we are composed, mathematics is to some lesser extent created in addition to being discovered. This makes it "more complete"than bare natural interactions and communications, which include the infinity of deductive theories *Skolem's theorem* versus the finite concrete perceptions themselves.

Gödel's theorems are a step away from the position that there is no theory of all theories. When we want to get to some distant truth, we need a lot of lies, and they are weak material, and by the time they arrive, the ship is sinking. I paraphrase the incompleteness theorems. It is even easier to understand the impossibility of a "*theory of everything*" using Russell's paradox of sets (Russell's paradox), which says that there is no set of all sets.

Namely, let there be a form of a universal set

$$Y = \{X \mid X \notin X\},$$

the set of all sets that are not members of themselves. This is actually a comfortable requirement because sets that are members of themselves are rare, but if such a Y turns out to be contradictory, then the idea of the set of all sets is especially contradictory. So, if Y is not a member of itself, then it precisely meets the condition of being a member of itself. This is an impossible situation, a contradiction that proves that there is no "set of all sets," the supposed universal set.

Russell vividly described the idea of this proof with the help of a village where there lives a barber who shaves all those who do not shave themselves. However, the barber is a local, so if he does not shave himself, then he should shave himself. This is also an impossible situation whose contradiction proves the impossibility of its existence. Also interesting is his reduction of God's omnipotence to a contradiction by asking, "Can God create a stone so large and heavy that he himself cannot lift it?"

These are slightly different from the simpler contradiction in the statement "I lie," which now reveals to us again that we can lie because we have vitality, that is, such a greater amount of options in relation to dead matter, which the commodity of living beings allows us. Its simplicity makes it easier to see the picture of the island of finitude, to which the natural phenomena we perceive with our senses are condemned, from the infinite sea of lies nad other obstacles that the multitude of such islands separate. The world of these liars and the truth is a story larger than the one offered to us by classical science, and the new theory of information will tell it.

1.3.4 Influence

Why are truths, in the narrow sense natural laws, more attractive to certain physical structures? What are the kinds of compulsions that drive them to enforce the rules? This rhetorical question is the guiding theme of the following topic.

The foundations of physical reality are based on the laws of conservation. From the symmetry section, we saw that when something lasts, it then has a past that is

growing and which is also counted in the totality, and because of which the present is thinned. The second is the *force of probability*, which equally works to reduce the uncertainty of the present, but by tending to translate states into more probable and less informative ones. Its thinning of the present is a reduction in the volume of events, their density, and number per unit of time, which we can understand as *slowing down of time*. This leads us to a radical demand: to turn the “thinning out” of the two above processes into a common statement: systems tend to a slower flow of time.

When we notice that naked natural phenomena are short-sighted and local, this statement does not seem as incredible as it may seem at first glance. Einstein himself stated in connection with his general theory of relativity that mass tends to place where time flows more slowly. By descending to lower atomic shells, the electron is freed from part of the energy $E = mc^2$, part of the mass m, or part of the frequency $\nu = mc^2/h$. This is a reduction in oscillations $\nu = 1/\tau$, and the extension of their duration τ we now understand as a slowing down of playfulness, so that the transition to a lower potential, or the tendency to a smaller effect, looks like a descent into a slower flow of time.

The equivalence of physical information and action would lead to significantly less diversity if there were no attraction of slower time. Already from the definition of a physical *action*, $S = \Delta E \cdot \Delta t$, we recognize periods $\tau = \Delta t$ of oscillations, say of a photon of energy $E = h\nu$ and frequency $\nu = 1/\tau$. Although this action is constant and quantized, it means that some such packets of particles-waves may be more attractive simply because they are then less "playful."

The same applies to molecules that oscillate less and less as they cool and thus lose information. Note that this is inconsistent with the classical understanding of entropy, which is officially equal to or at least proportional to information, while according to my theory it is the opposite. The gas in the container cools because the molecules transfer their playfulness, or information, to the surrounding cooler molecules on the walls of the container. This playfulness moves outwards as the entropy in the container increases. What is called “mess” are the shards of glass from a broken glass, which someone eventually needs to clean up, and which can also be considered a better arrangement of crystals, separate and impersonal, less informative.

I also see the effects of general tendencies towards greater probability, less information, and less action in laziness. It is a way of reducing our body’s communication as well as a tendency towards the familiar, not changing, or inertia. The desire for consistency would overcome the conflict within us and the resulting unpleasant feeling that arises from holding two opposing beleifs, values, or attitudes. These mental discomforts, the desire for inaction, the need for order, as well as all tendencies towards less information emission, come from the principle minimalism.

The need for *submission* to someone, or to a community, is at its root the same aspiration of principled minimalism. It is a way that frees a person from the excess information (amount of options) that he or she otherwise has in relation to the dead matter of which he or she is made, but he or she cannot do it just like

that, because all the surrounding substance would want the same. In the case of success, the “reward” comes with feelings of *security*, or efficiency, which are actually announcements of a reduction in freedoms, or of those that are partly well invested.

To clarify the latter, the freedom invested in *efficiency*, I will compare it to the swing of a pickaxe to drive deeper. Excess information gives living beings greater potential power to express themselves more qualitatively. In the appendix Letter Frequency, in translation *letter frequency*, simulations show how better text has less information. In short, you can find the same programs on my website (Kodovi, Nalazi slova), which analyzed samples of texts by Andrić, Selimović, Dostoevsky, or Fitzgerald, with the same results. A better text is more focused on a topic, less scattered about irrelevant things, and less informative in the sense of Shannon information. Even otherwise low-level subjects can have such depth, so even a text with little information does not have to be of great vitality.

Like the need for submission, the need for *domination* springs from a principled minimalism. If it is a dominant person, regardless of some privileges and additional freedoms, he has even greater “royal” restrictions. They relate to the lack of privacy and comfort in small things of most subjects, in the excess of obligations and routines necessary for the ruler. The freedom of superiors obtained from subordinates is a bitter cake, many will say, but we eat it gladly precisely for fear of excess uncertainties lurking outside. It seems absurd, but the same pattern is found in the need of electrons for a lower potential (Quantum Mechanics, [10], 1.4.6 Potential obstacles) and the desires of a powerful person for even more power.

If it is a community, whether an ant colony or an organized *state*, individuals have surrendered their freedoms to it in return for the expectation of receiving security and efficiency. From these expectations and from justice, the state gains permanence, which means reality, and, from the very freedoms of its subjects, *vitality*. Therefore, not only living beings are “vital,” but also their organizations, and some synergies and emergences. In the way that we are just seeing, reality also has *justice* to the extent that it contributes to duration.

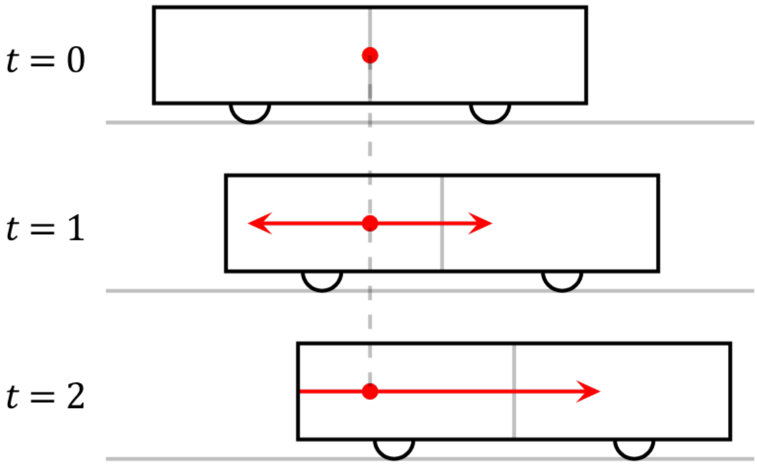
Let us now return from these digressions to the duration of the action, $S = \Delta E \cdot \Delta t$, to an increasingly longer and longer interval, $\Delta t \to \infty$, and an ever-decreasing energy $\Delta E \to 0$, so that their product does not change, $\Delta S =$ const. We get something timeless without energy, but so attractive to subjects who can interact with such S that it irresistibly reminds us of *natural law*. As the joke goes, if it looks like a duck, walks like a duck, and quacks like a duck, then it is a duck. Jokes aside, but we consider the fabric of space, time, and matter to be information and uncertainty to be its essence, so impossible and certain events are those with probabilities 0 and 1, respectively. Consequently, therefore, to the acceptance of infinity, the "laws"are also universal ($\Delta t \to \infty$) actions ($S = \Delta E \cdot \Delta t$).

We like order when we have too much freedom, and we don’t like prison when we have too little. Excess scares and challenges us, and deficiency creates a feeling of anxiety and rebellion, so we usually evaluate people as more attractive when they tell the truth than when they lie because fear overcomes anxiety. Let me remind you, a lie gives inanimate nature an excess of options and makes it vital, and

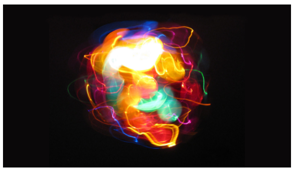
many of us do not know in advance whether these options are really real or just fictitious. Moreover, the fear of change stems from the inconstancy of the liar as well as his inability to unravel realities, because usually a good manipulator is not a good implementer. More on the latter later.

1.3.5 Simultaneity

Einstein (1905) discovered the relativity of *simultaneity* using a thought experiment of a man in a train and on an embankment, using the *principles of relativity* of motion and the independence of the speed of light from the speed of the source. A man in a train lights a match while standing exactly in the middle of the compartment, and the light from the phosphor reaches two opposite walls of the carriage simultaneously for him. However, as in the sketch, the train is moving to the right relative to the relative observer on the embankment, and the light from the match reaches the back of the carriage first and only then the front. Supposedly, the same simultaneous events for the first are not simultaneous for the second observer.



Let us now imagine this situation with the light stream reversed. For a passenger on a train, two simultaneous flashes of light come from opposite walls of the compartment and strike a thin sheet of paper standing upright exactly in the middle. The pressure of the opposing pulses of light does not knock the sheet of paper over. But for an observer on the embankment, the jet of light from the front of the carriage reaches the sheet of paper first and the sheet falls.



As a reminder, one of the recent achievements of the LHC is photons colliding head-on (A collision of light). Therefore, this simultaneous head-on collision of two flashes of light should not be considered impossible in an experiment. Also, mutual extinctions and transparencies of multiple incident light waves are achieved (Frontiers), here more significant due to theoretical treatments of similar events.

One can imagine that the falling of a note in the carriage triggers some chain reactions, even the release of poisonous gas that would kill the passengers like Schrödinger's cat. When the train reaches the station, and the man from the embankment appears there, the question arises: what will he see, and what will the other passengers think about this event?

From the point of view of information theory, given the implied additional *dimensions of time* (Dimensions) of this, the answers to that question are expected and interesting. First, the passengers on the train and the observer on the embankment at the station will report that they saw the same thing: the ticket that fell and the eventual consequences. This is not because the other event, the simultaneity and the maintenance of the ticket upright, did not "really" happen, but because they are not the same passengers.

Namely, standing at the same station, the passengers from the train and the man from the embankment are in the same present and have equal simultaneity. All other relatively moving systems could have had their own, some of them irrelevant simultaneity with the associated consequences. There are as many directions of movement as there are spatial dimensions (3-D), and so there are as many possible temporal ones, which is a total of six within such parallel multiverses. This is a rough estimate of the results from the given link—derived from the topological inductive definition of dimension and the postulation of objective uncertainty.

Second, there is the absence of the effect of the falling of the ticket and the possible consequences for the observer of simultaneous events. This is emphasized when all other observers are on the same side, at mutual rest at the station, but it becomes absurd when both the man in the compartment and the one on the embankment witness the (different) simultaneity of the same event, say, the lighting of a match in the middle of the compartment. Then we have "phantom action at a distance", in the words of Einstein (1935) when he discovered quantum entanglement, thereby attempting to deny the correctness of the mathematics of quantum mechanics.

I hope that you are already familiar with the decay of a particle with the emission of two entangled photons, so I am just adding this new explanation of the phenomenon. Such two photons originate from the same place at the same time, and as they move away, they form simultaneous events, a quantum entanglement. Their total spin is zero, but different (± 1) and objectively randomly distributed in those two photons.

If an experimenter on one side of these simultaneous photons catches one and measures its spin, say +1, then on the other side the possibly measured second photon would have to have a spin of -1. Also, conversely, if the found spin of the first is -1, then the found spin of the second would have to be +1. This is a "miracle"of quantum entanglement, that photons can be very far apart at the moment of measurement, and that second photon behaves as if it had received information from the first faster than light, and even instantaneously, about how to align.

From the example of the match, we now know that instantaneous events do not exchange actions or information. The first measurement on them is completely uncertain, and when it takes over the information from the photon by the measuring device, the uncertainty is removed from the entire coupled system so that the certainty of the second photon remains. They are like a pair of gloves hidden, one in each of the distant boxes, when we know exactly what is in the second after opening the first (Einstein's example).

It is interesting how these old descriptions of quantum entanglement, derived only from the formulas of quantum mechanics and since then otherwise incomprehensible, inexplicable, now fit so easily into this information theory. That is certainly a reason not to dismiss it lightly. I will mention that my first intention with it, postulating objective randomness in the early 1980s, was to seek a contradiction to disprove randomness and to prove the then tacit premise of *determinism* (the belief that all actions are conditioned before they are performed).

The same explanation of this theory is valid in the case of more recent discoveries of quantum entanglement, for example, when working with rubidium atoms whose spin is entangled with the spin of individual photons (Atom-Photon Entanglement). The Rb atom is placed in an optical dipole trap and excited when it decays from a short-lived upper state and emits a photon. By diverging, the atom and the photon form simultaneous events and a quantum entanglement.

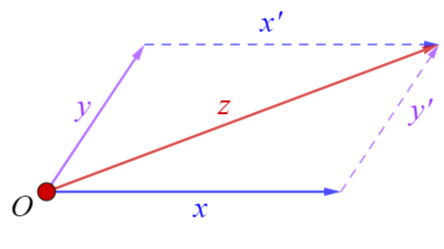
Glava 2

Functional

Unless otherwise stated, the default labels $X, Y, ...$ are vector spaces. The set of linear mappings $\mathcal{L}(X, Y)$, of vectors from the first space, usually labeled $x \in X$, to the second space $y \in Y$, is also a vector space. We denote the set of scalars by Φ, or by $\mathbb{R}$, or by $\mathbb{C}$ when we know that its elements are real, i.e., complex numbers. When $X = Y$, the mapping $A \in \mathcal{L}(X) = \mathcal{L}(X, X)$ is called a (linear) operator. A linear mapping of vectors onto scalars is called a functional, and the space of linear functionals on X, labeled $X^\dagger$, or $\mathcal{L}(X, \Phi)$, is the dual space of the vector space X.

2.1 Vectors

In mathematics, vectors are ordered sequences, linear mappings; they are also polynomials, solutions to differential equations, and the like. In physics, vectors are interpreted as forces, impulses, velocities, and then quantum states and hence quantum processes, and further here we interpret them as physical states and processes in general.



In the sketch on the left, there is a column at the fulcrum (O) where forces can be measured. Two forces, x and y, act on it in such a way that the measurement on the column shows the resulting pulling force, which we write as $z = x + y$. That is why we say that the vectors add up along the *parallelogram of forces*. That picture also shows the equality of the vectors on opposite sides of the parallelogram, $x' = x$ and $y' = y$, which means that *translation* (parallel displacement) does not change the vector. Hence the definition of vectors in general.

2.1.1 Vector Multiplication

The vector space over the field Φ is an additive Abelian group $X = \{x, y, z, ...\}$ in which multiplication with elements from Φ is defined such that for every pair

$x \in X$ and $\lambda \in \Phi$ there exists $\lambda x \in X$ where:

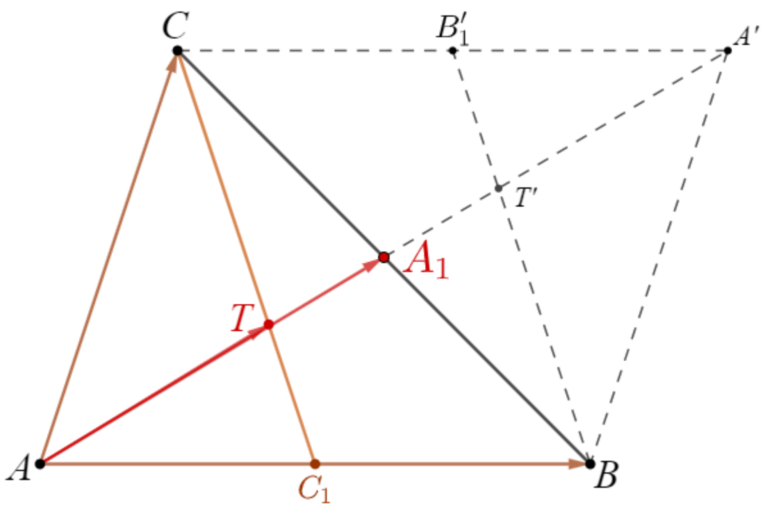
1. $\alpha(x+y) = \alpha x + \alpha y$
2. $(\alpha+\beta)x = \alpha x + \beta x$
3. $\alpha(\beta x) = (\alpha\beta)x$
4. $1 \cdot x = x$

for all vectors $x, y \in X$ and all scalars $\alpha, \beta \in \Phi$.

This is the general definition of a vector. It contains the multiplication of vectors and scalars, and it is derived from the addition of vectors and the mutual multiplication of scalars. Namely, $x+x = 2x$, or $2x+2x+2x = 6x$, then $3x+2ix = (3+2i)x$, and even $5x = (3+2i)(3-2i)x$.

Let's now look at a problem about the *centroid* of a triangle (the line joining the vertex to the midpoint of the opposite side) and the *center of gravity* (the intersection of the centroids) of a triangle. We use arrows above the letters to indicate vectors of oriented lines.

Example 23. *Let A_1 be the midpoint of side BC of triangle ABC. Show that $\overrightarrow{AA_1} = \frac{1}{2}(\overrightarrow{AB} + \overrightarrow{AC})$ and $\overrightarrow{AT} = \frac{2}{3}\overrightarrow{AA_1}$, where point T is the centroid of the triangle.*



Proof. In the image of triangle ABC, AA_1 and CC_1 are medians, so their intersection T is the centroid of the triangle. Through the line BC, the triangle is axisymmetrically mapped into triangle $A'BC$, the centroid T into the centroid T', and the center C_1 into the center B_1'. We know that an axisymmetric mapping does not change the distances between points (it is an isometry), maps parallel lines into parallel lines, and maps segments into proportional segments. Hence $AT = TT' = T'A'$, and A_1 is the midpoint of the line TT'. Therefore, $\overrightarrow{AA_1} = \frac{1}{2}(\overrightarrow{AB} + \overrightarrow{AC})$ and $\overrightarrow{AT} = \frac{2}{3}\overrightarrow{AA_1}$. □

The scalar product of vectors $\vec{a}$ and $\vec{b}$ is the scalar defined by $\vec{a}\cdot\vec{b} = ab\cos\angle(\vec{a},\vec{b})$, where $a = |\vec{a}|$ and $b = |\vec{b}|$ are the intensities of these vectors, $\vec{a}$ and $\vec{b}$, and $\angle(\vec{a},\vec{b})$ is the oriented angle between them. The provided link lists many useful properties of scalar multiplication that are assumed to be familiar to us.

You will find that the intensity of a vector is defined by $a^2 = |\mathbf{a}|^2 = \mathbf{a}\cdot\mathbf{a}$. We also write vectors in bold because they do not only refer to oriented lines. Then: $\mathbf{a}\cdot\mathbf{a} \ge 0$, $\mathbf{a}\cdot\mathbf{a} = 0$ if[1] $\mathbf{a} = 0$, and $\mathbf{a}\cdot\mathbf{b} = \mathbf{b}\cdot\mathbf{a}$. The following example proves the *identity of a parallelogram*, an important property of the scalar product.

Example 24. *Let's show that* $|\mathbf{a}+\mathbf{b}|^2 + |\mathbf{a}-\mathbf{b}|^2 = 2(|\mathbf{a}|^2 + |\mathbf{b}|^2)$.

Proof.

$$|\mathbf{a}+\mathbf{b}|^2 + |\mathbf{a}-\mathbf{b}|^2 = (\mathbf{a}+\mathbf{b})\cdot(\mathbf{a}+\mathbf{b}) + (\mathbf{a}-\mathbf{b})\cdot(\mathbf{a}-\mathbf{b}) =$$
$$= (\mathbf{a}\cdot\mathbf{a}+\mathbf{a}\cdot\mathbf{b}+\mathbf{b}\cdot\mathbf{a}+\mathbf{b}\cdot\mathbf{b}) + (\mathbf{a}\cdot\mathbf{a}-\mathbf{a}\cdot\mathbf{b}-\mathbf{b}\cdot\mathbf{a}+\mathbf{b}\cdot\mathbf{b})$$
$$= 2(\mathbf{a}\cdot\mathbf{a}+\mathbf{b}\cdot\mathbf{b}) = 2(|\mathbf{a}|^2+|\mathbf{b}|^2),$$

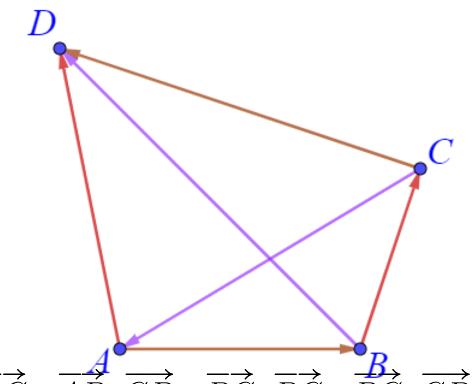
and that was what was supposed to be shown. □

Because of the way it is written, the scalar product is also called the *dot product*. However, it is also written using Dirac's *bra-ket* notation, $\langle x, y\rangle = x\cdot y$. A comma or vertical bar is placed between the vectors. However, this notation is more useful than a dot between when the components of the vector are complex numbers and when it is necessary to note that $\langle x|$ and $|x\rangle$ are, respectively, covariant and contravariant forms of vectors of the same property x.

Example 25. *Let* A, B, C, D *be four points in space. Then:*

$$\langle\overrightarrow{AB},\overrightarrow{CD}\rangle + \langle\overrightarrow{BC},\overrightarrow{AD}\rangle + \langle\overrightarrow{CA},\overrightarrow{BD}\rangle = 0.$$

The figure shows four points, A, B, C, D. Three of each, say A, B, C, are in a plane to which D does not necessarily belong. Express the desired vectors by $\overrightarrow{AB}$, $\overrightarrow{BC}$, and $\overrightarrow{CD}$, so:



$$\langle\overrightarrow{AB},\overrightarrow{CD}\rangle + \langle\overrightarrow{BC},\overrightarrow{AD}\rangle + \langle\overrightarrow{CA},\overrightarrow{BD}\rangle =$$
$$= \overrightarrow{AB}\cdot\overrightarrow{CD} + \overrightarrow{BC}\cdot(\overrightarrow{AB}+\overrightarrow{BC}+\overrightarrow{CD}) +$$
$$+(-\overrightarrow{AB}-\overrightarrow{BC})\cdot(\overrightarrow{BC}+\overrightarrow{CD}).$$

The scalar multiplication of these further is:

$$\overrightarrow{AB}\cdot\overrightarrow{CD} + \overrightarrow{BC}\cdot\overrightarrow{AB} + \overrightarrow{BC}\cdot\overrightarrow{BC} + \overrightarrow{BC}\cdot\overrightarrow{CD} - \overrightarrow{AB}\cdot\overrightarrow{BC} - \overrightarrow{AB}\cdot\overrightarrow{CD} - \overrightarrow{BC}\cdot\overrightarrow{BC} - \overrightarrow{BC}\cdot\overrightarrow{CD}$$

and that is zero.

Pseudo scalar product of vectors $\vec{a}$ and $\vec{b}$ is the scalar $[\vec{a},\vec{b}] = ab\sin\angle(\vec{a},\vec{b})$, where $a = |\vec{a}|$ and $b = |\vec{b}|$ are the intensities of these vectors, and $\angle(\vec{a},\vec{b})$ is the oriented angle between them. Behind the provided link is a commutator, because its value is exactly this.

[1] if — if and only if

Example 26. *Let's show that:*

$$[\lambda\mathbf{a},\mathbf{b}] = \lambda[\mathbf{a},\mathbf{b}], \quad [\mathbf{a},\mathbf{b}+\mathbf{c}] = [\mathbf{a},\mathbf{b}] + [\mathbf{a},\mathbf{c}],$$

for arbitrary vectors $\mathbf{a}$*,* $\mathbf{b}$ *and* $\mathbf{c}$ *and the scalar* λ*.*

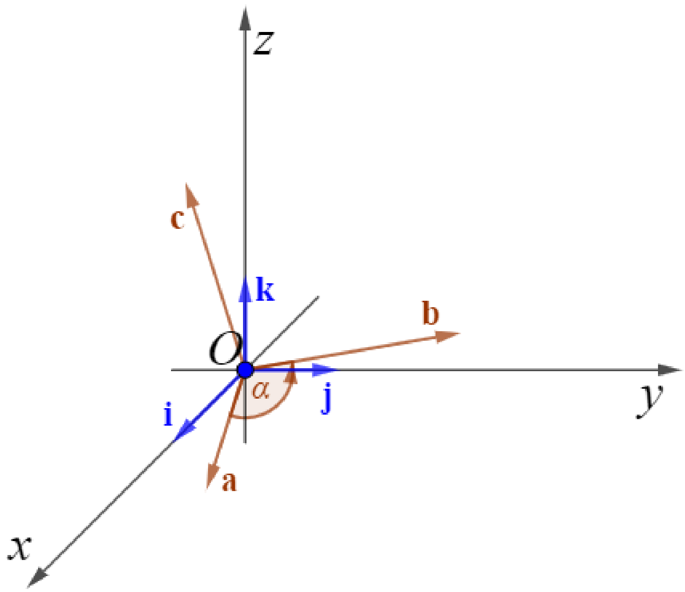
Proof.

$$[\lambda\mathbf{a},\mathbf{b}] = \lambda a_x b_y - \lambda a_y b_x = \lambda(a_x b_y - a_y b_x) = \lambda[\mathbf{a},\mathbf{b}],$$

$$[\mathbf{a},\mathbf{b}+\mathbf{c}] = a_x(b_y + c_y) - a_y(b_x + c_x) = (a_x b_y - a_y b_x) + (a_x c_y - a_y c_x) = [\mathbf{a},\mathbf{b}] + [\mathbf{a},\mathbf{c}].$$

That's what was supposed to be shown. □

Vector product of vectors $\mathbf{a}$ and $\mathbf{b}$ is the vector $\mathbf{c} = \mathbf{a}\times\mathbf{b}$ perpendicular to the parallelogram spanned by the factors $\mathbf{a}$ and $\mathbf{b}$ of the right-hand rule and of magnitude equal to its area, $c = ab\sin\alpha$, where $\alpha = \angle(\mathbf{a},\mathbf{b})$ is the angle between the factors. Unlike scalar multiplication, which we call the *inner product*, we call the cross product the *outer product* of vectors. In writing, we distinguish it from the cross between the first by the dot between the second.



The figure shows the right Cartesian rectangular coordinate system $(Oxyz)$ where the direction of the z-axis (applicate) is indicated by the thumb of the right hand when the fingers point in the direction from the x-axis (abscissa) to the y-axis (ordinate). The *orthogonal* (perpendicular) unit vectors of the $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ axes are drawn in blue. Therefore, we have:

$$\mathbf{i}\times\mathbf{j} = \mathbf{k}, \quad \mathbf{j}\times\mathbf{k} = \mathbf{i}, \quad \mathbf{k}\times\mathbf{i} = \mathbf{j}.$$

However, $\mathbf{a}\times\mathbf{b} = -\mathbf{b}\times\mathbf{a}$, so by changing the order of the multiplications, we change the sign of the result. When vectors are parallel, they do not open a parallelogram, and their cross product is zero. Hence the definition of *parallelism* $\mathbf{a}\|\mathbf{b}$ if $\mathbf{a}\times\mathbf{b} = 0$.

Compare this to dot multiplication when $\mathbf{a}\cdot\mathbf{a} = a^2$ and $\mathbf{a}\cdot\mathbf{b} = 0$ if $\mathbf{a}\perp\mathbf{b}$. The dot product of vectors is zero if and only if the vectors are mutually *orthogonal*, which means that the zero vector is orthogonal to every vector. The zero vector is also parallel to every vector. Furthermore, unlike the *law of commutation* $\mathbf{a}\cdot\mathbf{b}\cdot\mathbf{a}$ in inner multiplication, the holds for the outer product *law of alternation*, $\mathbf{a}\times\mathbf{b} = -\mathbf{b}\times\mathbf{a}$. It is easy to check the *laws of distribution*:

$$\mathbf{a}\cdot(\mathbf{b}+\mathbf{c}) = \mathbf{a}\cdot\mathbf{b}+\mathbf{a}\cdot\mathbf{c}, \quad \mathbf{a}\times(\mathbf{b}+\mathbf{c}) = \mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c}$$

and I leave that to the reader.

If we introduce new notations $\mathbf{e}_\nu$ for orthogonals $(\mathbf{i},\mathbf{j},\mathbf{k})$, where the indices are the axis labels, or better still the ordinal numbers $\nu\in\{1,2,\dots,n\}$, then the ideas of inner and outer multiplication can be easily generalized to a rectangular system of $n\in\mathbb{N}$ dimensions. Then we use the so-called ε-*symbol*, or *Levi-Civita symbol* (for $n = 3$):

$$\varepsilon_{ijk} = \begin{cases} 1, & \text{even permutation } ijk \\ -1, & \text{odd permutation } ijk \\ 0, & \text{at least two indices equal} \end{cases}$$

Vector products can now be written in the form

$$\mathbf{e}_j\times\mathbf{e}_k = \varepsilon_{ijk}\mathbf{e}_i$$

So, say, $\mathbf{e}_1\times\mathbf{e}_2 = \varepsilon_{112}\mathbf{e}_1+\varepsilon_{212}\mathbf{e}_2+\varepsilon_{312}\mathbf{e}_3 = \mathbf{e}_3$ for $j = 1$ and $k = 2$. In that system, the vectors are:

$$\mathbf{a} = a_1\mathbf{e}_1+a_2\mathbf{e}_2+a_3\mathbf{e}_3, \quad \mathbf{b} = b_1\mathbf{e}_1+b_2\mathbf{e}_2+b_3\mathbf{e}_3,$$

from where we find the dot product:

$$\mathbf{a}\cdot\mathbf{b} = a_1b_1+a_2b_2+a_3b_3$$

and the vector product:

$$\mathbf{a}\times\mathbf{b} = \varepsilon_{ijk}a_jb_k\mathbf{e}_i = (a_2b_3-a_3b_2)\mathbf{e}_1+(a_3b_1-a_1b_3)\mathbf{e}_2+(a_1b_2-a_2b_1)\mathbf{e}_3.$$

The same, written using determinants:

$$\mathbf{a}\times\mathbf{b} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}. \tag{2.1}$$

Coordinate reflection, $x_i' = -x_i$ $(i = 1,2,3)$, is a change in the sign of the direction, the orientation of the coordinate axes. In such a transformation, vectors that by their nature have a certain orientation (force, position) also change sign, $\mathbf{a}' = -\mathbf{a}$. They are called *real vectors*, or "polar vectors." However, there are also vectors that do not change when the coordinates are reflected, for which $\mathbf{a}' = \mathbf{a}$, which is the moment of force. Such vectors are called *pseudo vectors*, or "axial vectors".

Task 27. *Let's write the expression* $\mathbf{a}\times\mathbf{b}+\mathbf{b}\times\mathbf{c}+\mathbf{c}\times\mathbf{a}$ *in the form of a vector product.*

Solution. Using the laws of distribution and alternation we find:

$$\mathbf{a}\times\mathbf{b}+\mathbf{b}\times\mathbf{c}+\mathbf{c}\times\mathbf{a}=(\mathbf{a}-\mathbf{c})\times\mathbf{b}+\mathbf{c}\times\mathbf{a}-\mathbf{a}\times\mathbf{a}=$$

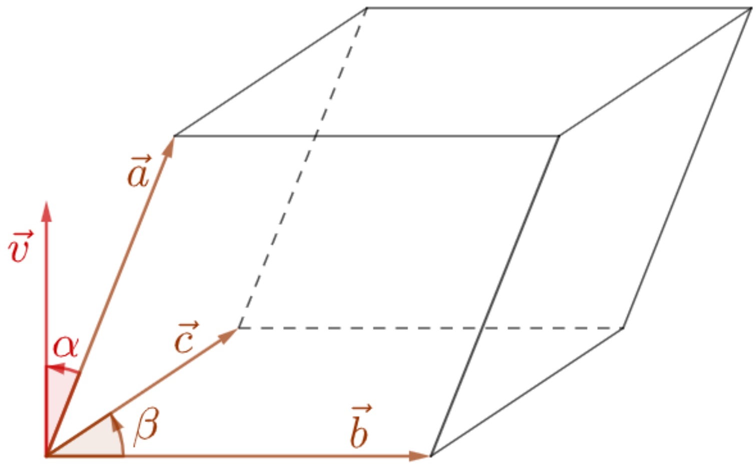
$$=(\mathbf{a}-\mathbf{c})\times\mathbf{b}-(\mathbf{a}-\mathbf{c})\times\mathbf{a}=(\mathbf{a}-\mathbf{c})\times(\mathbf{b}-\mathbf{c}).$$

So, $(\mathbf{a}-\mathbf{c})\times(\mathbf{b}-\mathbf{c})$ is the desired result. □

The mixed product of vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ is the scalar product of the first vector with the cross product of the next two, $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$. When we look at the parallelepiped that spans these three vectors and take the parallelogram of the last two as the base, their vector product, $\mathbf{v}=\mathbf{b}\times\mathbf{c}$, is in the direction of the height of the parallelogram and the intensity of its area, $v=bc\sin\angle(\mathbf{b},\mathbf{c})$, of the base of the parallelepiped. Therefore:

$$\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})=\mathbf{a}\cdot\mathbf{v}=V[\mathbf{a},\mathbf{b},\mathbf{c}].$$

The mixed vector product is equal to the oriented volume of the parallelepiped spanning the three vectors.



This can be seen in the figure. The area of the base parallelogram is $|\mathbf{v}|=|\mathbf{b}\times\mathbf{c}|=bc\sin\beta$, and the height of the parallelepiped is $\mathbf{a}\cdot\mathbf{u}=a\cos\alpha$, where $\mathbf{u}=\mathbf{v}/|\mathbf{v}|$ is the unit height vector. The height multiplied by the base gives the volume of the parallelepiped, $\mathbf{a}\cdot\mathbf{u}|\mathbf{v}|=V$. Using the determinant notation (2.1), we now find:

$$\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})=V[\mathbf{a},\mathbf{b},\mathbf{c}]=\begin{vmatrix}a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{vmatrix}. \tag{2.2}$$

From this equality, referring to the properties of determinants, we easily derive the basic properties of the mixed product:

$$\mathbf{b}\cdot(\mathbf{b}\times\mathbf{c})=\mathbf{b}\cdot(\mathbf{c}\times\mathbf{a})=\mathbf{c}\cdot(\mathbf{a}\times\mathbf{b}).$$

The mixed product of vectors does not change under cyclic permutation of its factors.

Double vector product , $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$, is the vector product of the vector $\mathbf{a}$ and the vector product of two vectors $\mathbf{b} \times \mathbf{c}$. By the definition of a vector product, this vector is normal to $\mathbf{b} \times \mathbf{c}$. The double vector product lies in the plane formed by the vectors $\mathbf{b}$ and $\mathbf{c}$, so it is of the form:

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \alpha \mathbf{b} + \beta \mathbf{c}.$$

On the other hand, developing into components, we find:

$$[\mathbf{a} \times (\mathbf{b} \times \mathbf{c})]_1 = a_2(\mathbf{b} \times \mathbf{c})_3 - a_3(\mathbf{b} \times \mathbf{c})_2 = a_2(b_1c_2 - b_2c_1) - a_3(b_3c_1 - b_1c_3).$$

Add and subtract $a_1b_1c_1$ to this expression, and we get:

$$[\mathbf{a} \times (\mathbf{b} \times \mathbf{c})]_1 = b_1(a_1c_1 + a_2c_2 + a_3c_3) - c_1(a_1b_1 + a_2b_2 + a_3b_3) =$$
$$= b_1(\mathbf{a} \cdot \mathbf{c}) - c_1(\mathbf{a} \cdot \mathbf{b}).$$

Repeating this with indices 2 and 3, we find analogous expressions for the other components and assemble the final result:

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b}). \tag{2.3}$$

Now let's look at some examples of how this is used.

Example 28. *Let us show using vectors that the area of a parallelogram constructed over its diagonals is twice its area.*

Solution. The vector expression of this is $(\mathbf{a} - \mathbf{b}) \times (\mathbf{a} + \mathbf{b}) = 2(\mathbf{a} \times \mathbf{b})$, where the vectors $\mathbf{a}$ and $\mathbf{b}$ are the sides of the parallelogram, and their difference and sum are its diagonals. However:

$$(\mathbf{a} - \mathbf{b}) \times (\mathbf{a} + \mathbf{b}) = \mathbf{a} \times \mathbf{a} + \mathbf{a} \times \mathbf{b} - \mathbf{b} \times \mathbf{a} - \mathbf{b} \times \mathbf{b} = 2(\mathbf{a} \times \mathbf{b}).$$

This means that the vector equality is true. □

Example 29. *Let's prove Jacobi's identity:*

$$(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} + (\mathbf{b} \times \mathbf{c}) \times \mathbf{a} + (\mathbf{c} \times \mathbf{a}) \times \mathbf{b} = 0.$$

Proof. Using (2.3) we find:

$$(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - \mathbf{a}(\mathbf{b} \cdot \mathbf{c}),$$
$$(\mathbf{b} \times \mathbf{c}) \times \mathbf{a} = \mathbf{c}(\mathbf{a} \cdot \mathbf{b}) - \mathbf{b}(\mathbf{a} \cdot \mathbf{c}),$$
$$(\mathbf{c} \times \mathbf{a}) \times \mathbf{b} = \mathbf{a}(\mathbf{b} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b}).$$

Adding these together yields the required Jacobi identity. □

Example 30. *Let's show that:*

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) + (\mathbf{b} \times \mathbf{c}) \cdot (\mathbf{a} \times \mathbf{d}) + (\mathbf{c} \times \mathbf{a}) \cdot (\mathbf{b} \times \mathbf{d}) = 0.$$

Proof. The left side of the given equality is equal to:

$$[(\mathbf{a} \times \mathbf{b}) \times \mathbf{c}] \cdot \mathbf{d} + [(\mathbf{b} \times \mathbf{c}) \times \mathbf{a}] \cdot \mathbf{d} + [(\mathbf{c} \times \mathbf{a}) \times \mathbf{b}] \cdot \mathbf{d} =$$
$$= [(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} + (\mathbf{b} \times \mathbf{c}) \times \mathbf{a} + (\mathbf{c} \times \mathbf{a}) \times \mathbf{b}] \cdot \mathbf{d} = 0,$$

because the zero vector in the square brackets (Jacobi's identity). □

2.1.2 Linear dependence

In short, *linear independence* is characterized by the absence of redundancy. To elaborate a little, let's imagine that we have an original writer and a plagiarist (of a scientific paper, a school assignment, music, or the like). They may even think the same way, anyway; whatever idea the first has, the other has it too. Then one of them is irrelevant, redundant, or linearly dependent. It is nothing more than redundant information. In contrast, the idea of "linear dependence"would be the accumulation of the same, or the same values.

Let's imagine that we are talking about three types of fruit, say apples, strawberries, and cherries. A kilogram of each fruit is represented by one of the three basis vectors $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$. No amount of apples can replace a certain amount of strawberries, or a certain amount of strawberries a certain amount of cherries, or the like. We further demonstrate linear dependence with these fruits. Thus:

$$\mathbf{a} = 2\mathbf{e}_1 + \mathbf{e}_2 + 3\mathbf{e}_3$$

represents a basket with two kilograms of apples, one kilogram of strawberries, and three kilograms of cherries. Let's say there are two more baskets:

$$\mathbf{b} = \mathbf{e}_1 + 3\mathbf{e}_2 + 2\mathbf{e}_3, \quad \mathbf{c} = 3\mathbf{e}_1 + 2\mathbf{e}_2 + \mathbf{e}_3$$

and we say we want the first two, the second three, and the third four. When asked how many fruit trees we have in total and which ones, we calculate:

$$\mathbf{v} = 2\mathbf{a} + 3\mathbf{b} + 4\mathbf{c} =$$

$$= 2(2\mathbf{e}_1 + \mathbf{e}_2 + 3\mathbf{e}_3) + 3(\mathbf{e}_1 + 3\mathbf{e}_2 + 2\mathbf{e}_3) + 4(3\mathbf{e}_1 + 2\mathbf{e}_2 + \mathbf{e}_3)$$
$$= (2\cdot 2 + 3 + 4\cdot 3)\mathbf{e}_1 + (2 + 3\cdot 3 + 4\cdot 2)\mathbf{e}_2 + (2\cdot 3 + 3\cdot 2 + 4)\mathbf{e}_3$$
$$= 19\mathbf{e}_1 + 19\mathbf{e}_2 + 16\mathbf{e}_3.$$

So, we have 19 kilograms of apples and strawberries and 16 kilograms of cherries. But, the vector $\mathbf{v}$ is in a given way, by that equation, dependent on the basis vectors $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$. Or, in the way $\mathbf{v} = 2\mathbf{a} + 3\mathbf{b} + 4\mathbf{c}$ it is dependent on their linear combinations, the baskets $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ of fruit.

Example 31. *In a pile* $\mathbf{v}$ *we have six kilograms of each of three fruits: apples, strawberries, and cherries. How many baskets of* $\mathbf{a}$*,* $\mathbf{b}$*, and* $\mathbf{c}$ *are there?*

Solution. If there were α, β, and γ kilograms of the fruits listed in the pile (now $\alpha = \beta = \gamma = 6$, so we can guess the solution without calculating), the problem would be more interesting, and we would calculate:

$$x\mathbf{a} + y\mathbf{b} + z\mathbf{c} = \alpha\mathbf{e}_1 + \beta\mathbf{e}_2 + \gamma\mathbf{e}_3,$$

$$x(2\mathbf{e}_1+\mathbf{e}_2+3\mathbf{e}_3)+y(\mathbf{e}_1+3\mathbf{e}_2+2\mathbf{e}_3)+z(3\mathbf{e}_1+2\mathbf{e}_2+\mathbf{e}_3)=\alpha\mathbf{e}_1+\beta\mathbf{e}_2+\gamma\mathbf{e}_3,$$

$$(2x+y+3z)\mathbf{e}_1+(x+3y+2z)\mathbf{e}_2+(3x+2y+z)\mathbf{e}_3=\alpha\mathbf{e}_1+\beta\mathbf{e}_2+\gamma\mathbf{e}_3.$$

Due to the impossibility of replacing apples with strawberries or cherries, we say that due to the linear independence of the basis vectors e_i ($i=1,2,3$), we separate this vector equation into three ordinary ones:

$$\begin{cases} 2x+y+3z=\alpha \\ x+3y+2z=\beta \\ 3x+2y+z=\gamma \end{cases}$$

This is a system of linear equations that is further solved in standard ways. In the given case $\alpha=\beta=\gamma=6$, the solution is $x=y=z=1$. We have one of each of the three baskets of fruit. □

Depending on the given number of fruits in the pile, it is possible to get half a basket, or less, or even negative baskets in the solution to make the result even. These are linear equations that write *linear combinations* values and express their linear dependecies. In general, a set of vectors is linearly dependent if one of them can be written as a linear combination of the others. If this is not possible, the set of vectors is linearly independent.

Each vector oriented along a set gives a specific, distinct direction in space. A set of vectors $\{\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_n\}$ is linearly independent if the equation:

$$\alpha_1\mathbf{v}_1+\alpha_2\mathbf{v}_2+...+\alpha_n\mathbf{v}_n=0$$

has only the trivial solution $\alpha_1=\alpha_2=...=\alpha_n=0$. If the set of vectors forms an orthogonal set (every pair of vectors is orthogonal), then they are linearly independent.

Example 32. *Let us show that the following three polynomials* $\{1+x,2x+x^2,2+x-3x^2\}$ *are linearly independent vectors.*

Solution. We assume that the linear combination of the given vectors vanishes:

$$\alpha_1(1+x)+\alpha_2(2x+x^2)+\alpha_3(2+x-3x^2)=0,$$

$$(\alpha_1+2\alpha_3)+(\alpha_1+2\alpha_2+\alpha_3)x+(\alpha_2-3\alpha_3)x^2=0,$$

$$\alpha_1+2\alpha_3=0,\quad \alpha_1+2\alpha_2+\alpha_3=0,\quad \alpha_2-3\alpha_3=0.$$

The only solution is trivial: $\alpha_1=\alpha_2=\alpha_3=0$, which means that the given three polynomials form a linearly independent set of vectors. □

Independent vectors form a basis of a space *number of dimensions* equal to the number of those vectors. Of particular interest are *orthogonal bases*, those whose scalar products of all pairs of vectors are zero. For them to be orthogonal, they would also have to be of unit intensity (unit norm).

Example 33. *Show that the following three vectors are linearly independent:*

$$v_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

and form an orthogonal base.

Solution. We are looking for all solutions to the equation:

$$\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 = 0,$$

$$\alpha_1 \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \alpha_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

$$\begin{pmatrix} \alpha_1 + \alpha_2 \\ -\alpha_1 + \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

$$\begin{cases} \alpha_1 + \alpha_2 = 0 \\ -\alpha_1 + \alpha_2 = 0 \\ \alpha_3 = 0 \end{cases}$$

the only solution is trivial, $\alpha_1 = \alpha_2 = \alpha_3 = 0$, which means that these vectors are linearly independent. By direct multiplication, we also determine the orthogonality of these vectors, $v_1 \cdot v_2 = v_1 \cdot v_3 = v_2 \cdot v_3 = 0$. □

When we have a basis space, then each vector of that space can be written as a linear combination of the basis vectors. This is the case in this example:

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \frac{a-b}{2} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + \frac{a+b}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + c \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

for arbitrary scalars a, b, c. Matrices are also vectors, and the following second-order quadratic matrices are all the more interesting because the square of each, $\sigma_\mu^2 = I$, $\mu \in \{x, y, z\}$, is the identity matrix (I). These are *Pauli matrices*, known from quantum physics:

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.4}$$

This imaginary unit ($i^2 = -1$) will always be easily distinguished from the index of the same symbol depending on the use in the text, or it will be specially indicated. When multiplied by it, Pauli matrices become *quaternions*:

$$q_x = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad q_y = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad q_z = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}. \tag{2.5}$$

By direct matrix multiplication, we easily verify $q_\mu^2 = -I$, $\mu \in \{x, y, z\}$.

Pauli matrices are unit squares, but we cannot speak of orthogonality, so they are not orthogonal. However, including the unit, they are linearly independent and form a basis of a 4-dimensional vector space. Namely:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \frac{a_{11}+a_{22}}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{a_{21}+a_{12}}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$-\frac{i(a_{21}-a_{12})}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} + \frac{a_{11}-a_{22}}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

This means that every second-order square matix is a linear combination of Pauli matrices, and with the identity matrix, they thus represent a "real"4-dimensional matrix vector space. If we want a "pseudo-real"vector space, say like relativistic space-time, we can include some of the quaternions as a matrix replacement for the imaginary unit. Matrices are generally not commutative, and this gives a special quality to matrix vector spaces.

2.1.3 Vector Norm

Norm is a function defined on a vector space that associates to each vector a measure of its length. There is a close connection between the norm and the scalar product, since any scalar product can be used to induce a norm on its space; for example, $\|v\| = \sqrt{v \cdot v}$ is the norm of a vector v in Euclidean space. The norm of a vector $u, v \in X$ is defined by the following four axioms.

1. Non-negativity: $\|v\| \ge 0$.
2. Specificity: $\|v\| = 0$ ако и само ако $v = 0$.
3. Homogeneity: $\|\alpha v\| = |\alpha|\|v\|$, где $\alpha \in \Phi$ је скалар.
4. Triangle inequality: $\|u + v\| \le \|u\| + \|v\|$.

These properties are quite intuitive. The norm measures the length of a vector, and (1) it says that it is not a negative number, (2) it is zero if the vector has length zero. As we increase the vector, its norm changes by the same amount, (3) it says, and (4) the length of a vector is the shortest distance between its ends. Let's look at some examples of norms of discrete vectors, $x = (\xi_1, \xi_2, ..., \xi_n) \in X$, of a finite-dimensional vector space X of dimension $n \in \mathbb{N}$.

Manhattan norm is another name for the L_1 norm:

$$\|x\|_1 = |\xi_1| + |\xi_2| + ... + |\xi_n|.$$

It is the sum of the absolute values of the coordinates ($\xi_k \in \Phi$) of a vector. It is clear that it satisfies the above axioms.

Euclidean norm is the name for the L_2 norm:

$$\|x\| = \sqrt{|\xi_1|^2 + |\xi_2|^2 + \dots + |\xi_n|^2}.$$

It is the root of the sum of the squares of the absolute values of the coordinates. The above axioms are derived from the L_2 norm, so the subscript in its notation $\|x\|_2$ is often implied.

Max norm is the name for the L_∞ norm:

$$\|x\|_\infty = \max_{1\le k\le n} |\xi_k|.$$

It is the largest of the absolute values of ehr coordinates of the vector. Obviously, axioms 1, 2 and 3 are fulfilled, and 4 can be seen from:

$$\|x+y\|_\infty = \max_{1\le k\le n}\{|\xi_k + \eta_k|\} \le \max_{1\le i\le n}\{|\xi_i|\} + \max_{1\le j\le n}\{|\eta_j|\} = \|x\|_\infty + \|y\|_\infty$$

where the vector $y = (\eta_1, \eta_2, \dots, \eta_n) \in X$.

It is less well known that all of the above are just special cases of p-*norm* ($p \ge 1$), of the space of the designation L_p:

$$\|x\|_p = (|\xi_1|^p + |\xi_2|^p + \dots + |\xi_n|^p)^{1/p}.$$

This is the p-th root of the sum of the p-th powers ($p \ge 1$) of the absolute values of the coordinates of the vector. The first three axioms are obviously true for this norm, and the triangle inequality follows from Minkowski's inequality. On the page Terence Tao you will find interesting views and evidence regarding this norm.

Example 34. *Let us prove that the triangle inequality holds for the p-norm:*

$$\|x+y\|_p \le \|x\|_p + \|y\|_p$$

with vectors $x = (\xi_1, \xi_2, \dots, \xi_n)$ and $y = (\eta_1, \eta_2, \dots, \eta_n)$ and parameter $1 \le p < \infty$.

Proof. Because of homogeneity, we can reduce the problem to the case of inequality for $\|x\|_p = 1-\lambda$ and $\|y\|_p = \lambda$, where $0 < \lambda < 1$, with the limiting cases, $\lambda \in \{0,1\}$ being clearly true. Let $x' = x/(1-\lambda)$ and $y' = y/\lambda$ and use convexity:

$$|(1-\lambda)\xi'_k + \lambda\eta'_k|^p \le (1-\lambda)|\xi'_k|^p + \lambda|\eta'_k|^p.$$

Adding these inequalities for all $k = 1, 2, \dots, n$ we obtain:

$$(\|(1-\lambda)x' + \lambda y'\|_p)^p \le 1,$$

and hence, by returning the reduction, the required triangle inequality follows. □

A similar idea of reducition due to homogeneity is used (Distances III, 1st paragraph) to prove that the norm $\|x\|_p$ on the same vector $x = (\xi_1, \xi_2, \dots, \xi_n)$ is smaller the larger the parameter $p \ge 1$ is. That blog also discusses q-norms that are dual to p-norms and vice versa when $1/p + 1/q = 1$.

On page 43, the *scalar product* of "ordinary" vectors is defined, and the "parallelogram identity" is proven. There is a note about the importance of such a property of this vector multiplication, because it is considered that every norm obtained from the scalar, or inner product, should satisfy the same identity. However, p-norms for $p \neq 2$ do not satisfy the parallelogram identity. Therefore, every inner product defined on a vector space can lead to a norm on that space, but the converse is not true. The norms L_p, for $p \neq 2$, are an example of such that no inner product generates. Read more about these norms in the section on Banach space.

The *Matrix norm* $\|A\|_p$ is the largest, or *supremum* (least upper bound of the set) value $\|Au\|_p$ that the matrix A can achieve with unit vectors $u = x/\|x\|$, $x \neq 0$:

$$\|A\|_p = \sup_{\|u\|=1} \|Au\|_p.$$

It also satisfies the above axioms of vector norm. The essential difference between matrix norm and vector norm is multiplication, and hence matrices are especially useful if they are also submultiplicative, if $\|AB\| \le \|A\|\|B\|$.

Let the given matrix be:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

When $p = 1$, the norm of the matrix becomes the maximum of the sum of the absolute vavlues of its columns, and when $p = \infty$, the norm of the matrix is the maximum of the sum of the absolute values of its rows:

$$\|A\|_1 = \max_{1\le j\le n} \sum_i^m |a_{ij}|, \qquad \|A\|_\infty = \max_{1\le i\le m} \sum_{j=1}^n |a_{ij}|.$$

This is easily verified by multiplying the matrix by the unit vectors $u = (v_1, v_2, \dots, v_n)$, where $\|u\|_1 = \sum_k |v_k| = 1$, while $\|u\|_\infty = \max_k |v_k| = 1$.

Example 35. *Given a matrix*

$$A = \begin{pmatrix} 1 & 2 \\ -5 & 3 \end{pmatrix}.$$

Let us check that its $p = 1$ and $p = \infty$ norms are $\|A\|_1 = 6$ and $\|A\|_\infty = 8$, respectively.

Solution. In general:

$$Au = \begin{pmatrix} 1 & 2 \\ -5 & 3 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} v_1 + 2v_2 \\ -5v_1 + 3v_2 \end{pmatrix}.$$

For $p = 1$, since $|v_1|+|v_2| = 1$, we are looking for the maximum of $|v_1+2v_2|+|-5v_1+3v_2|$. And this is obvious when $v_1 = 1$ and $v_2 = 0$, so $\|A\|_1 = 6$. The second norm of the vector $u = (-1, 1)$, however, gives the (largest) of $|v_1 + 2v_2|$ and $|-5v_1 + 3v_2|$ for the latter, so $\|A\|_\infty = 8$. □

Matrix norms for other cases, $1 < p < \infty$, can be very different. However, one of the most common is the *Frobenius norm* . It is the square root of the sum of the squares of all the elements in the matrix. It gives an impression of the total "size" of the matrix:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2}.$$

It is also equal to the square root of the matrix trace, simply because the trace of the matrix does not change with a change in the base, and then it is the same for a diagonal matrix, which makes the value of this expression equal to the previous one:

$$\|A\|_F = \sqrt{\mathrm{Tr}(AA^\dagger)}.$$

Here $A^\dagger = (A^*)^\top$ adjoint matrix, i.e. conjugate transpose of the matrix A. The Frobenius norm is included in the vector p-norms for $p = 2$, so it is also written in the form $\|A\|_2$. It can be shown that it is equal to the product of the largest eigenvalue of the matrix and the root of the product of the adjoint matrices:

$$\|A\|_2^2 = \sup_{x\in X}\frac{\|Ax\|_2^2}{\|x\|_2^2} = \sup_{x\in X}\frac{\langle x, A^\dagger Ax\rangle}{\langle x, x\rangle} = |\lambda_{\max}|^2(A^\dagger A).$$

Here $\langle\cdot,\cdot\rangle$ is the scalar product of the 2-norm of the space X, and $Ax = \lambda x$ is the eigenequation of the matrix A.

The corresponding p-norm of integrable functions $x(t)$ on the interval $t \in (a, b)$ is $L^p(a, b)$:

$$\|x\|_p = \left(\int_a^b |x(t)|^p dt\right)^{\frac{1}{p}},$$

where also the parameter $p \ge 1$. Again, for $p = 1$ $\|x\|_1$ is the ordinary integral of the absolute values of the function; for $p = \infty$ $\|x\|_\infty$ is the supremum of the function $x(t)$ for $t \in (a, b)$ and the space labels $M(a, b)$. You can find some more details about norms on my page (Metrics).

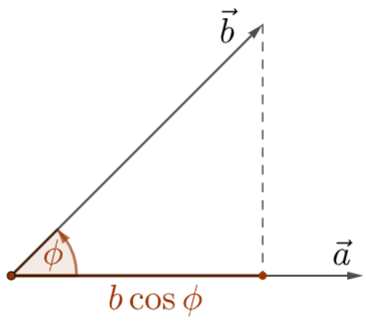
2.1.4 Scalar Product

Scalar product of vectors is a binary operation of two vectors $\vec{a}, \vec{b} \in X$ whose result is a scalar $s \in \Phi$. The figure shows the scalar product of oriented lines.

The angle between vectors $\vec{a}$ and $\vec{b}$ is $\phi = \angle(\vec{a}, \vec{b})$. The projection of vector $\vec{b}$ onto vector $\vec{a}$ is $b\cos\phi$, where $b = |\vec{b}|$ is the intensity, i.e., 2-norm, of vector $\vec{b}$. This multiplied by the intensity of vector $a = |\vec{a}|$ is:

$$s = \vec{a}\cdot\vec{b} = ab\cos\phi$$

This is the dot product of the given vectors. It is clearly commutative, $\vec{a}\cdot\vec{b} = \vec{b}\cdot\vec{a}$, as well as distributive, $\vec{a}\cdot(\vec{b}+\vec{c}) = \vec{a}\cdot\vec{b} + \vec{a}\cdot\vec{c}$.



We can look at vectors a little more broadly, say, in a space with more than two dimensions. Then $\vec{a} = (\alpha_1, \alpha_2, ..., \alpha_n)$ and $\vec{b} = (\beta_1, \beta_2, ..., \beta_n)$ in some Cartesian rectangular coordinate system of dimension $n = 1, 2, 3, ...$, and because $\cos 90° = 0$, we find that:

$$\vec{a} \cdot \vec{b} = \alpha_1\beta_1 + \alpha_2\beta_2 + ... + \alpha_n\beta_n.$$

No matter how many dimensions n, the previous picture remains the same, because two such vectors can always be reduced to one plane to span a parallelogram.

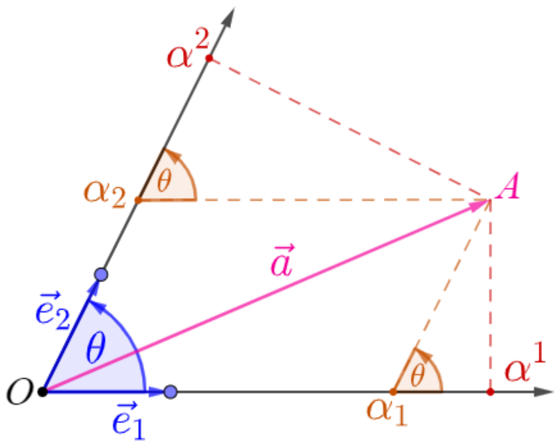
In the case of a skew coordinate system (Variance) we have two important types of vector representation. *Covariant coordinates* of a vector are parallel to the $\vec{e}_k$, where $k = 1, 2, ..., n$, unit vectors of the x_k axis. We usually write them with subscripts, $\vec{a} = (\alpha_1, \alpha_2, ..., \alpha_n)$. The other *contravariant* coordinates are perpendicular to the axes and are written with superscripts, $\vec{b} = (\beta^1, \beta^2, ..., \beta^n)$. Their scalar product is:

$$\vec{a} \cdot \vec{b} = \alpha_1\beta^1 + \alpha_2\beta^2 + ... + \alpha_n\beta^n.$$

The square of the intensity of a single such vector:

$$|\vec{a}|^2 = \alpha_1\alpha^1 + \alpha_2\alpha^2 + ... + \alpha_n\alpha^n$$

does not change with a change in the diagonal system.



Indeed, in the figure we see:

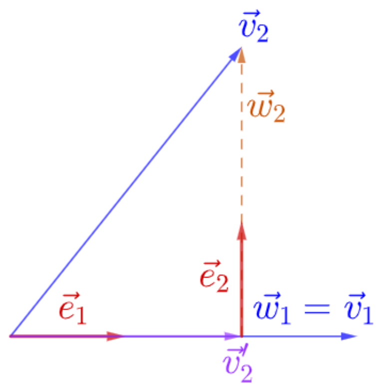
$$OA^2 = (\alpha_1)^2 + (\alpha_2)^2 + 2\alpha_1\alpha_2\cos\theta =$$

$$= \alpha_1(\alpha^1 - \alpha_2\cos\theta) + \alpha_2(\alpha^2 - \alpha_1\cos\theta) + 2\alpha_1\alpha_2\cos\theta,$$

$$OA^2 = \alpha_1\alpha^1 + \alpha_2\alpha^2.$$

The cosine theorem (triangle $O\alpha_1 A$) and the legs of small triangles ($A\alpha_1\alpha^1$ and $A\alpha^2\alpha_2$) were used. The result is the *invariant* $OA = |\vec{a}|$ to the change in angle $\theta = \angle(\vec{e}_1, \vec{e}_2)$. The intensity of the vector does not change with the change in the angle of the oblique system.

Quantum physics uses complex numbers, the set $\mathbb{C}$, for scalars Φ of the vector space X. Then, the transition from the covariant to the contravariant coordinates requires their conjugation and transposition, which is otherwise called adjoint. Thus, $x^\dagger = (x^*)^\top$ is the adjoint vector $x \in X$. It is customary to write covariant vectors as matrices of the type and contravariant vectors as column matrices. All these properties are clearly written in Dirac's *bra-ket* brackets, and this notation for the scalar product $\langle a, b\rangle$ leaves the additional possibility that a and b can be the names of a series of properties needed for the experiment. Quantum mechanics works almost exclusively with orthonormal bases, unit perpendicular basis vectors.



The *Gram-Schmidt procedure* of orthonormalizing a set of vectors $\{v_1, v_2, ..., v_n\}$ is a repetition of taking orthogonal projections and normalizing them. A typical step of this procedure is shown in the figure on the left.

The first vector $\vec{w}_1 = \vec{v}_1$ is taken, and the unit vector $\vec{e}_1 = \vec{w}_1/|\vec{w}_1|$ is formed. The next datum $\vec{v}_2' = \langle \vec{v}_2, \vec{e}_1\rangle \vec{e}_1$ is projected onto it. The difference $\vec{w}_2 = \vec{v}_2 - \vec{v}_2'$ is perpendicular to the previous datum, $\vec{w}_2 \perp \vec{w}_1$, and it is then also normalized, $\vec{e}_2 = \vec{w}_2/|\vec{w}_2|$. This is repeated, and a sequence of perpendicular vectors $\vec{w}_k$ and a sequence of their normalized $\vec{e}_k = \vec{w}_k/|\vec{w}_k|$ are obtained.

For example, let us consider three non-commutative but linearly independent vectors:

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \quad \vec{v}_2 = \begin{pmatrix} -2 \\ -1 \\ 1 \end{pmatrix}, \quad \vec{v}_3 = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}.$$

Since $\langle v_1, v_3\rangle = \vec{v}_1 \cdot \vec{v}_3 = 1\cdot 3 + 0 \cdot 2 + 2 \cdot 1 = 5 \ne 0$, they are not orthogonal. From the difference of $\vec{v}_1$ and $\vec{v}_2 + \vec{v}_3$, we see that they are not linearly dependent either. We find the orthogonals:

$$\vec{w}_1 = \vec{v}_1, \quad \vec{w}_2 = \vec{v}_2 - \frac{\vec{w}_1 \cdot \vec{v}_2}{|\vec{w}_1|^2}\vec{w}_1, \quad \vec{w}_3 = \vec{v}_3 - \frac{\vec{w}_1 \cdot \vec{v}_3}{|\vec{w}_1|^2} - \frac{\vec{w}_2 \cdot \vec{v}_3}{|\vec{w}_2|^2}\vec{w}_2,$$

$$\vec{w}_1 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \quad \vec{w}_2 = \begin{pmatrix} -2 \\ -1 \\ 1 \end{pmatrix}, \quad \vec{w}_3 = \frac{1}{6}\begin{pmatrix} -2 \\ 5 \\ 1 \end{pmatrix}.$$

It is easy to check that $\vec{w}_1 \cdot \vec{w}_2 = \vec{w}_1 \cdot \vec{w}_3 = \vec{w}_2 \cdot \vec{w}_3 = 0$, which means that these vectors are orthogonal. We normalize them, $\vec{e}_k = \vec{w}_k/|\vec{w}_k|$, with k = 1, 2, 3, to obtain the orthonormal ones:

$$\vec{e}_1 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \quad \vec{e}_2 = \frac{1}{\sqrt{6}}\begin{pmatrix} -2 \\ -1 \\ 1 \end{pmatrix}, \quad \vec{e}_3 = \frac{1}{\sqrt{30}}\begin{pmatrix} -2 \\ 5 \\ 1 \end{pmatrix}.$$

The orthonormal basis $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ spans the same space as the set of initial vectors $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$.

The procedure is similar when the coefficients of the vector are not only real numbers ($\mathbb{R}$), but also complex numbers ($\mathbb{C}$). Then $\langle x, y\rangle = x^* \cdot y = \xi_1^*\eta_1 + \xi_2^*\eta_2 + \dots + \xi_n^*\eta_n$, where the prime factors are like the conjugate transposes of the primes.

2.1.5 Information of Perception

The basic idea of information theory was that the weaving of space, time, and matter is a type of information and that the essence of these is uncertainty. This leads us to the very bottom of physics, to quantum mechanics and quantum states as an interpretation of quantum vectors. Their lowest representations would be superpositions, or probability distributions in the manner of probability waves and the Born's law, and then climbing into the macro world, we would get their linear combinations. Like the atoms and molecules invisible to us that make up the visible world of chemical substances, uncertainty creates the macro world.

However, information is a specific "substance", because it is by its nature a kind of freedom, a quantity of options. It goes beyond the *principle of least action*, which is the alpha and omega of theoretical physics, at least in the way of "living beings" that can wander outside the framework of these minima. In the book "Information Perception" [7] it is noted that the "amount of activity", let us denote it by the number a, is proportional to the "freedoms", the size s, and is inversely proportional to the "constraints", the number b. In other words, the freedom $s = ab$ is proportional to both activities and constraints.

There was success in such an interpretation of physical phenomena, as you can see from the following book, "Space-Time" [8] which I wrote almost simultaneously with the first, among other things in extending the aforementioned "freedoms" to their various forms and linear combinations, the simplest of which is:

$$s = s_1 + s_2 + \dots + s_n = a_1b_1 + a_2b_2 + \dots + a_nb_2.$$

This is a form of the scalar product of vectors, similar to Shannon information:

$$S = -p_1 \log p_1 - p_2 \log p_2 - \dots - p_n \log p_n.$$

Of course, the story of information with such properties is just beginning.

When it comes to uncertainty, there is also its repulsive force that forces more likely events into more frequent outcomes. As if hiding its luxuries from reality, nature tends to avoid participation. Understanding this leads us to minimalism, the natural tendency for things to happen that are more likely and less informative. And the extreme of this is the principle of least action – which is the backbone of physics. Therefore, the topics of physics are a subset of those of informatics, and the question of expansion arises.

For the purpose of expansion, I used the somewhat clumsy working title vitality, in the expectation that something else besides *living beings* might have it. So far, I have also observed it (vitality) in the organizations of living beings, and only in hints perhaps in synergy, or emergence. Minimalism drives everything living and non-living to get rid of excess options, but precisely because everyone would do the

same, for which the laws of conservation also apply, there is a crowd on the field. In that small pond with a lot of crocodiles, the individual manages to transfer his unwanted burden to the collective. I am paraphrasing, of course, what in life we call laziness, devotion, or the death drive, and in still life inertia.

Despite the law of conservation, it is not possible for only what has been to be, even if repackaged to the point of being difficult to recognize, because that would contradict the principle of uncertainty. Nature is always moving towards something, even if it is completely different from what it has ever been. The universe is constantly being created, in addition to being recreated. In addition to phenomena that last and are counted in the total amount of information, it also has phenomena that do not last and may never appear again. For example, in addition to reality, we have fictions, in addition to truths, we have lies.

Our fictions and reality float on a vast pool of the unknown. They appear in various forms, states, which we translate into numbers using the above formula. We map vectors into scalars, which means that this formula is functional. There is a theorem of linear algebra, Riesz's theorem, which says that for every mapping f of a vector like x from the space X to the scalars Φ, there is one and only one vector $y \in X$ that this mapping defines with $f(x) = \langle x, y \rangle$. In other words, the subject (y) is *unique* in the way it depicts and perceives its environment ($x \in X$).

When $Y, Z \subset X$ are complementary (3.3. Proposition), then to each $x \in X$ there uniquely corresponds some $y \in Y$ and $z \in Z$ such that $x = y + z$. Hence, the complement of known states *complementary* (separate parts that make up a whole) will not violate the uniqueness of the subject of perception.

Using the aforementioned formula for information, therefore, we have unique concrete participants in the cosmos, but also multiple copies of parts of abstract truths (the Löwenheim-Skolem theorem). We can imagine these particularities as bulbs of countless abstract truths, of which only the finite ones are available to us, but they cannot be the only ones (hard truths). Otherwise, they would violate the principle of uncertainty. Randomness is an objective phenomenon; it is not only ignorance that can be revealed to us by information but is also a much broader "ignorance."

Multiplicity is a prerequisite for concealing and revealing secrets, and in that sense it is the other side of uncertainty, or the essence of information. It was the topic of the book of the same name [9], but it is a much broader phenomenon than that, as evidenced by Gödel's and similar impossibility theorems. Therefore, the information of perception, $s = \langle a, b \rangle$, no matter how broad a mathematical concept it may seem to us, should be seen as a bottleneck through which a small part of those secrets is revealed to us. Reality leaks to us through those numbers (s), barely dripping, and we cannot even consider anything that cannot be perceived as real.

To be more precise, the states x and y are separate only potential options, unrealized ghosts that can only be some reality in conjunction $\langle x, y \rangle$. The object A that can be observed (interact) with the object B_1, this one with B_2, ..., this one with B_n and finally that one with C, is reality for C. In other words, in order to have "reality," we need some (discrete, intermittent) continuity, duration. In the world of information, "real" is that for which some law of conservation applies. But, alas,

there is no singularity without uniqueness, so the realities that we observe are indeed always intermittent (discrete).

Note that this is a slightly different contribution to quantization than, say, the description of packages. It is important because of the seemingly absurd finding (2.7. Proposition) that the linear mapping $A : X \to Y$ is continuous if and only if it is bounded, in other words, because holding to finiteness is lame without insight into infinity.

2.2 Processes

States are representations of vectors (elements of $X, Y, \ldots$), and processes of linear mappings ($A : X \to Y$). However, mappings are also a type of vector, so they can also be interpreted as states, states of processes. *Lineal* is otherwise the vector space of linear mappings between two given sets, $\mathcal{L}(X, Y)$, and a special lineal is X^* is the space of linear functionals on X, of linear mappings ($x^* \in X^*$) of vectors into scalars ($x^* : X \to \Phi$). The vector space X^* is the dual of the vector space X.

I repeat the notations here so that I don't have to define them all the time in the text. Also, there is a note (I quote) that the use of linear operators alone limits quantum theory very much, like, say, the decision to use only linear functions in areas of mathematics.

2.2.1 Hamiltonian

Hamiltonian is a fundamental concept in quantum physics. Strictly speaking, it is an operator that represents the total energy of a quantum system, including kinetic and potential. It provides a mathematical framework for describing the dynamics and evolution of a quantum state and is part of the *Schrödinger equation*:

$$i\hbar \frac{d|\psi\rangle}{dt} = \hat{H}|\psi\rangle. \tag{2.6}$$

The Hamiltonian $\hat{H}$ is a Hermitian operator on the state space of a quantum system. The eigenvalues of the Hamiltonian correspond to the possible energy levels of the system, and the associated eigenvectors are the corresponding stationary states.

Analogous to classical mechanics, the Hamiltonian is the sum of the kinetic and potential energy operators, $\hat{H} = \hat{T} + \hat{U}$, where the potential energy operator is a multiplication, $\hat{U} = U(r, t)$, and the kinetic energy and momentum operator:

$$\hat{T} = \frac{\hat{p} \cdot \hat{p}}{2m} = -\frac{\hbar^2}{2m}\nabla^2, \quad \hat{p} = -i\hbar\nabla$$

expressed using the nabla operator. This should be distinguished from the *Lagrangian*, $L = T - U$, which is the difference between kinetic and potential energy. We used the latter, via the Euler-Lagrange equations (Example 18 and Problem 19), to discover conics as charge trajectories under the action of central forces and to observe the decrease of these forces with the square of the distance.

In both cases, the momentum is $p = mv = \partial_v(mv^2 - U) = \partial_v L$. If we write this in generalized coordinates q, the generalized momentum will be $p = \partial_{\dot{q}} L$. Hence, from $H = mp\dot{q} - L = 2T - (T - U) = T + U$, the connection between the Lagrangian and the Hamiltonian is obtained. On the other hand, the work done by the force on the path is written as $\partial_q H = F$, and the force is also the change in momentum over time, $F = -\dot{p}$, so:

$$\dot{q} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial q}. \tag{2.7}$$

Example 36. *Let us show that the Hamiltonian of the particle is in spherical coordinates:*

$$\hat{H} = \frac{1}{2m}\left(p_r^2 + \frac{p_\theta^2}{r^2} + \frac{p_\phi^2}{r^2 \sin\theta}\right) + V(r, \theta, \phi).$$

Proof. Kinetic energy is of the form:

$$T = \frac{mv^2}{2} = \frac{m}{2}(v_r^2 + v_\theta^2 + v_\phi^2) = \frac{m}{2}\left[\dot{r}^2 + (r\dot{\theta})^2 + (r\sin\theta\dot{\phi})^2\right].$$

Momenta are the derivatives of kinetic energy with respect to velocities:

$$p_r = \frac{\partial T}{\partial \dot{r}} = m\dot{r}, \quad p_\theta = \frac{\partial T}{\partial \dot{\theta}} = mr^2\dot{\theta}, \quad p_\phi = \frac{\partial T}{\partial \dot{\phi}} = mr^2\sin^2\theta\dot{\phi}.$$

Solving these equations in terms of velocities, we find:

$$\dot{r} = \frac{p_r}{m}, \quad \dot{\theta} = \frac{p_\theta}{mr^2}, \quad \dot{\phi} = \frac{p_\phi}{mr^2\sin^2\theta},$$

$$\hat{H} = \frac{1}{2m}\left(p_r^2 + \frac{p_\theta^2}{r^2} + \frac{p_\phi^2}{r^2 \sin\theta}\right) + V(r, \theta, \phi),$$

which should have been shown. □

The velocities along the coordinates, changing with time, in this example are expressed as derivatives of the Hamiltonian with respect to the momenta:

$$\dot{r} = \frac{\partial H}{\partial p_r}, \quad \dot{\theta} = \frac{\partial H}{\partial p_\theta}, \quad \dot{\phi} = \frac{\partial H}{\partial p_\phi}.$$

This is, therefore, also an example of generalized coordinates of the canonical equations (2.7).

Let's take a brief look at all of the above and the Hamilton equation *harmonic oscillator*, that extremely important process for physics.

$$L = \frac{m\dot{x}^2}{2} - \frac{kx^2}{2}, \quad p = \frac{\partial L}{\partial \dot{x}} = m\dot{x},$$

$$H = \dot{x}p - L = \dot{x}p - \left(\frac{m\dot{x}^2}{2} - \frac{kx^2}{2}\right) = \frac{m\dot{x}}{2} + \frac{kx^2}{2},$$

$$\dot{x} = \frac{\partial H}{\partial p} = \frac{p}{m}, \quad \dot{p} = -\frac{\partial H}{\partial x} = -kx,$$
$$F = -kx.$$

With the spring's elastic coefficient, $k > 0$, the spring force, $F = -kx$, acts opposite and proportional to its extension along the abscissa. This is Hook's law. The momentum is too great, and the spring pulls back to the equilibrium position ($x = 0$), accelerating, increasing the kinetic energy, and due to inertia, the transfer to the other side occurs again, so the phenomenon goes again. It is a "process"that repeats itself, but it is also a štate"of oscillation, of alternating kinetic and potential energy.

Analogously, information is news that becomes stale immediately after its appearance. The vacant place of its disappearance is attractive, and a deficit occurs that overcomes the spontaneous resistance to the appearance of a new related announcement. Thus, a "process" of repetition continues, which we can also call the "state" of information oscillation.

Let's go back for a moment to the previous title (1.3.2 Vitality) and the simulation of obtaining šomething"from "nothing."There, two matrices A and B were attached, which disappeared with successive multiplication (repetition), but their mixed products and sums were preserved and even gave an isometry that we interpret with conservation laws.

Analogous to them are the *ladder operators* of quantum mechanics, respectively lowering $\hat{a}$ and raising $\hat{a}^\dagger$. The latter is the conjugate transpose (adjoint) of the former:

$$\hat{a}|n\rangle = \sqrt{n}|n-1\rangle, \quad \hat{a}^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle.$$

They shift, say, the n-th energy state lower or higher, which can also be the eigenstates of the Hamiltonian for a quantum harmonic oscillator.

Example 37. *Let us show that:*

$$(\hat{a}) = \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & \dots \\ 0 & 0 & \sqrt{2} & 0 & \dots \\ 0 & 0 & 0 & \sqrt{3} & \dots \\ 0 & 0 & 0 & 0 & \dots \\ \dots \end{pmatrix}, \quad (\hat{a}^\dagger) = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots \\ \sqrt{1} & 0 & 0 & 0 & \dots \\ 0 & \sqrt{2} & 0 & 0 & \dots \\ 0 & 0 & \sqrt{3} & 0 & \dots \\ \dots \end{pmatrix}$$

matrix representations of ladder operators of lowering and raising.

Proof. We put the "sandwich" operator in the brackets to obtain the coefficients of their matrices $(\hat{a})_{ij}$ and $(\hat{a}^\dagger)_{ij}$:

$$\langle i|\hat{a}|j\rangle = \sqrt{j}\langle i|j-1\rangle = \sqrt{j}\delta_{i,j-1}, \quad \langle i|\hat{a}^\dagger|j\rangle = \sqrt{j+1}\langle i|j+1\rangle = \sqrt{j+1}\delta_{i,j+1}.$$

The coefficients of the matrices are $(\hat{a})_{ij} = \langle i|\hat{a}|j\rangle$ and $(\hat{a}^\dagger)_{ij} = \langle i|\hat{a}^\dagger|j\rangle$. We wrote them using Kronecker's *delta symbols*, $\delta_{ij} = 1$ when $i = j$ and $\delta_{ij} = 0$ when $i \neq j$, hence the above matrix representations. $\square$

These operators are not commutative:

$$\hat{a}\hat{a}^\dagger|n\rangle = \hat{a}\sqrt{n+1}|n+1\rangle = (n+1)|n\rangle, \quad \hat{a}^\dagger\hat{a}|n\rangle = \hat{a}^\dagger\sqrt{n}|n-1\rangle = n|n\rangle,$$

$$\hat{a}\hat{a}^\dagger|n\rangle - \hat{a}^\dagger\hat{a}|n\rangle = |n\rangle,$$

$$[\hat{a}, \hat{a}^\dagger] = \hat{a}\hat{a}^\dagger - \hat{a}^\dagger\hat{a} = 1.$$

Their unitary "indeterminacy,"lowering and raising the number $n = 0, 1, 2, \dots$, is information (news, energy) that "oscillates."Only the commutator $[\hat{a}, \hat{a}^\dagger]$ is a physical state and is a Hermitian operator. Its second sum, $\hat{N} = \hat{a}^\dagger\hat{a}$, is called the *operator of numbers*, because its eigenvalue is also the number n. The operator is Hermitian, and such eigenvalues are real numbers; here $n \in \{0, \mathbb{Z}^+\}$.

In the case of *indeterminacy* of position and momentum, which also persists and makes the particle-wave state ψ sustainable, we have the constancy of them themselves. See the general case, task 12, and then the following example.

Example 38. *Let us show that:*

$$[\hat{x}, \hat{p}] = i\hbar$$

are the uncertainty relations for the momentum position operators.

Proof. The operators *position* and *momentum* are ($h = 2\pi\hbar$, $i^2 = -1$):

$$\hat{x}\psi = x\psi, \quad \hat{p}\psi = -i\hbar\frac{\partial}{\partial x}\psi = p\psi,$$

$$\hat{x}\hat{p}\psi = \hat{x}p\psi = xp\psi, \quad \hat{p}\hat{x}\psi = \hat{p}x\psi = -i\hbar\frac{\partial}{\partial x}(x\psi) = -i\hbar\psi + xp\psi,$$

$$\hat{x}\hat{p}\psi - \hat{p}\hat{x}\psi = i\hbar\psi,$$

$$[\hat{x}, \hat{p}] = \hat{x}\hat{p} - \hat{p}\hat{x} = i\hbar.$$

Hence $|\Delta x||\Delta p| \ge \hbar$ and the required relations. □

Thus, the commutator $[\hat{x}, \hat{p}]$ is not Hermitian, but the position and momentum operators are. The ladder operators expressed using the position and momentum operators are:

$$\hat{a} = \sqrt{\frac{m\omega}{2\hbar}}\left(\hat{x} + \frac{i}{m\omega}\hat{p}\right), \quad \hat{a}^\dagger = \sqrt{\frac{m\omega}{2\hbar}}\left(\hat{x} - \frac{i}{m\omega}\hat{p}\right),$$

where m is the mass of the particle, $\omega = 2\pi\nu$ is the circular frequency, and $\hbar\omega = h\nu$ is the energy.

Example 39. *Let's express the Hamiltonian operator using the number operator:*

$$\hat{H} = \hbar\omega\left(\hat{N} + \frac{1}{2}\right).$$

This is the Hamiltonian for the quantum harmonic oscillator.

Solution. The number operator is the product of ladder operators:

$$\hat{N} = \hat{a}^\dagger \hat{a} = \frac{m\omega}{2\hbar}\left(\hat{x} - \frac{i}{m\omega}\hat{p}\right)\left(\hat{x} + \frac{i}{m\omega}\hat{p}\right) =$$

$$= \frac{m\omega}{2\hbar}\left(\hat{x}^2 + \frac{\hat{p}^2}{m^2\omega^2} + \frac{1}{m\omega}[\hat{x},\hat{p}]\right)$$

$$= \frac{1}{\hbar\omega}\left(\frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega^2\hat{x}^2\right) - \frac{1}{2}$$

$$= \frac{1}{m\omega}(\hat{T} + \hat{U}) - \frac{1}{2}$$

$$= \frac{1}{m\omega}\hat{H} - \frac{1}{2},$$

$$\hat{H} = \hbar\omega\left(\hat{N} + \frac{1}{2}\right).$$

The Hamiltonian is obtained in the announced form. □

The Hamiltonian and number operators share the same eigenvector, the state $|n\rangle$, but they have different eigenvalues:

$$\hat{H}|n\rangle = E_n|n\rangle, \quad \hat{N}|n\rangle = n|n\rangle.$$

The observables E_n are energy levels, with the ground state energy E_0, and the number n is also real because both of these operators are Hermitian. Due to the constant 1/2, the two operators are also commutative, which is why they share the same eigenstate.

Example 40. *From the eigenvalues of the n operator of the number $\hat{N}$, we express:*

$$E_n = \hbar\omega\left(n + \frac{1}{2}\right)$$

the eigenvalues of the energy of the quantum harmonic oscillator.

Solution. From the above, we further calculate:

$$\hbar\omega\left(\hat{N} + \frac{1}{2}\right)|n\rangle = E_n|n\rangle,$$

$$\hat{N}|n\rangle + \frac{1}{2}|n\rangle = \frac{E_n}{\hbar\omega}|n\rangle,$$

$$\hat{N}|n\rangle = \left(\frac{E_n}{\hbar\omega} - \frac{1}{2}\right)|n\rangle = n|n\rangle,$$

$$n = \frac{E_n}{\hbar\omega} - \frac{1}{2}, \quad n + \frac{1}{2} = \frac{E_n}{\hbar\omega},$$

$$E_n = \hbar\omega\left(n + \frac{1}{2}\right) = (2n+1)\frac{\hbar\omega}{2},$$

$$E_0 = \frac{\hbar\omega}{2},\ E_1 = \frac{3\hbar\omega}{2},\ E_2 = \frac{5\hbar\omega}{2}, \dots,\ E_n = \frac{\hbar\omega(2n+1)}{2}, \dots$$

These are, in order, the ground state of energy, the first excited state, the second excited state, and so on, each subsequent one increased by a quantum of energy. $\Delta E = h\nu$. □

Example 41. *For a quantum harmonic oscillator, let us show that:*

$$(\hat{H}) = \hbar\omega \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 & \dots \\ 0 & \frac{3}{2} & 0 & 0 & \dots \\ 0 & 0 & \frac{5}{2} & 0 & \dots \\ 0 & 0 & 0 & \frac{7}{2} & \dots \\ \dots \end{pmatrix}$$

is the matrix representation of the Hamiltonian operator.

Proof. We start from the eigenvalue equation of the Hamiltonian:

$$\hat{H}|n\rangle = E_n|n\rangle.$$

This is actually the time-independent Schrödinger equation. Continuing with the results of the previous examples:

$$\hat{H} = \hbar\omega\left(\hat{N} + \frac{1}{2}\right), \quad E_n = \hbar\omega\left(n + \frac{1}{2}\right), \quad n = 0, 1, 2, \dots$$

Quantum mechanics works with orthonormal bases and for the matrix coefficient in such a representation we find:

$$(\hat{H})_{m,n} = \langle m|\hat{H}|n\rangle = \langle m|(E_n|n\rangle) = E_n\langle m|n\rangle = E_n\delta_{m,n},$$

where $\delta_{m,n}$ is the Kronecker delta symbol: it is equal to 1 when $m = n$ and 0 when $m \ne n$. From:

$$(\hat{H})_{m,n} = \hbar\omega\left(n + \frac{1}{2}\right)\delta_{m,n}$$

and in the order of row and column numbers $m, n = 0, 1, 2, \dots$ the above matrix follows. □

2.2.2 Position and momentum

The uncertainty relations for the position operators $\hat{x}$ and momentum $\hat{p}$ come from their noncommutativity. We derived them (example 38) for comparison with the similar *ladder operators* raising $\hat{a}^\dagger$ and lowering $\hat{a}$ which, in turn, reminded us of the earlier simulation of getting "something" from "nothing". This means that both "position" and "momentum" can be reduced to the same thing, and as we know:

$$\hat{x} = \sqrt{\frac{\hbar}{2m\omega}}(\hat{a}^\dagger + \hat{a}), \quad \hat{p} = i\sqrt{\frac{\hbar m\omega}{2}}(\hat{a}^\dagger - \hat{a}).$$

The formal derivation of these well-known expressions for a quantum harmonic oscillator is a secondary issue for now. Instead, let's start understanding position and momentum from a computational perspective.

We can compare position and momentum as a "state" and a "process" of the same phenomena, or as an "image" and a "film" (example 13) of the same representation. With a given memory limit, more of the displayed film comes at the expense of lower sharpness of individual images, and conversely, with higher image resolution, the amount of action decreases.

The development of energy over time seems more abstract to us than movement through space; it seems easier to make analogies with information. However, the transfer of energy over time and space is equally an informational phenomenon. Drawing the metric space of energy, actions, and the structure of physical reality, mathematics itself is completely indifferent; all these "drawings" are forms of the same impersonal vector spaces in which we "walk" as much as we "fantasize" about them. Basically, they are analogous starting vectors that build their linear combinations.

In other words, the progression from place to place from moment to moment, in many of the cases we describe, has a geometric and metric form. Nature is truly such (Skolem) that its forms are everywhere, but our concrete perceptions are unique and make it difficult to perceive that other side of the *cosmos*, its universalities. Countless copies of the fragments of the harmonic oscillator, which we get to know through the Hamiltonian, are also found in the operators of position and momentum. The oscillation of information that would like to disappear but cannot is fundamentally equivalent to walking through positions and velocities.

Hence the great connection of ladder operators with position and momentum processes. The familiar formalism remains:

$$(\hat{x}) = \sqrt{\frac{\hbar}{2m\omega}}\begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & \dots \\ \sqrt{1} & 0 & \sqrt{2} & 0 & \dots \\ 0 & \sqrt{2} & 0 & \sqrt{3} & \dots \\ 0 & 0 & \sqrt{3} & 0 & \dots \\ \dots \end{pmatrix}, \quad (\hat{p}) = i\sqrt{\frac{\hbar m\omega}{2}}\begin{pmatrix} 0 & -\sqrt{1} & 0 & 0 & \dots \\ \sqrt{1} & 0 & -\sqrt{2} & 0 & \dots \\ 0 & \sqrt{2} & 0 & -\sqrt{3} & \dots \\ 0 & 0 & \sqrt{3} & 0 & \dots \\ \dots \end{pmatrix}.$$

These two matrices, representing the position and momentum operators, are a simple sum of the above matrices of the ladder operators.

Using the vector norm, we can easily define the metric of the space, $d(x,y) = \|x - y\|$. In addition to those that we would get from the above examples (2.1.3 Vector Norm), we are also interested in the Hamming metric. Let us fix a natural number $n \in \mathbb{N}$ for the set X of binary strings (digits 0 or 1) of length n. The distance between two such strings, $d(x,y)$, is the number of positions in which these strings differ. For example, for $n = 3$, it will be $d(101,110) = 2$ and $d(011,001) = 1$. It is easy to verify that for this "distance" defined, the following axioms that define *metric* hold.

Definition 42. *Let X be an amorphous set whose elements are denoted by the letters $x, y, z, \dots$. If we coordinate to each ordered pair (x,y) in X a real number $d(x,y)$ that*

has the following properties:

1. $0 \le d(x,y) < +\infty$, *positivity,*
2. $d(x,y) = 0 \iff x = y$, *minimalism,*
3. $d(x,y) = d(y,x)$, *symmetry,*
4. $d(x,y) \le d(x,z) + d(z,y)$, *triangle inequality,*

we say that we have provided the set X with the metric d. A set X provided with a metric d is a metric space, its elements are points, and $d(x,y)$ is the distance between points x and y.

The Hamming metric can be constructed for much longer strings (large numbers $n \in \mathbb{N}$), when the digits are not only binary, and if they are also letters of the alphabet; otherwise, we would not be interested in it. It is clear that this is a mathematical, abstract phenomenon. Let us now imagine that the digits of these strings are variable. Let the probabilities of changing the digits of a single string form a distribution, such that the rightmost digits are more likely to change. We have obtained a simulation of (chaotic) motion. By giving higher probabilities to the initial values, we would have oscillations of points around some equilibrium points similar to the titration of atoms or molecules of a crystal.

I give this example not to amuse myself by elaborating on it now (that will be another time) but to inspire us in imagining abstractions that are fundamentally metric spaces. Like probability spaces, geometrically inconceivable phenomena can be reduced to geometric forms or to the equivalent of indeterminacy of position and momentum. Such are the dispersions σ_A and σ_B task 12.

Let us now transfer this model (Hemming metrics) to the genetic structure of living beings on Earth, in general of all species that have ever lived. Individuals are "points," and the greater the *genetic differences*, the greater the "distances" between them. Geneticists already use this metric model, although unaware of its mathematical basis, when determining the relationship from species to successive descendants. Faster genetic changes, in the steps of birth and death, mean less focus on the genes of the species. In other words, the species that changed faster through evolution had a greater dispersion of genes among its immediate descendants. Just like the dependence of two uncertainties, position on momentum.

When we have a metric, it makes sense to talk about lengths and areas and then about *commutators*. However, for commutators to represent uncertainty relations, as in the case of position and momentum (example 38), we need a "state" and a "process." Then, such uncertainty is also information for which the conservation law applies. In the amount of uncertainty, the values of the commutator $[x,y]$, the first variable x is the state, and the second y is the corresponding process. The higher the "resolution" of the first, the lower the second, and vice versa.

The *universe* is an interesting example of this. We believe that the Big Bang occurred 13.8 billion years ago, when the universe as we know it was born. It was like

a hot amorphous soup that expanded and crystallized, in terms of the appearance and detail of structure, with laws that evolved into more and more subtle and binding. At least that's what this information theory suggests. In this finding, the aforementioned switch speaks of a decreasing resolution of the beginning of the universe in favor of its increasing duration.

The commutator is also a *pseudo scalar* product of vectors, we now denote them by $\vec{x}$ and $\vec{y}$. It is a scalar $[\vec{x}, \vec{y}] = xy \sin\theta$, where $x = |\vec{x}|$ and $y = |\vec{y}|$ are the intensities of these vectors, and $\theta = \angle(\vec{x}, \vec{y})$ is the oriented angle between them. We interpret it as an indeterminacy, but geometrically its meaning is the area of the parallelogram spanned bu the given vectors. It increases with the angle θ. On the contrary, the ordinary *scalar product* of vectors, $\vec{x} \cdot \vec{y} = xy \cos\theta$, is a scalar that we interpret as "perception information" (2.1.5) whose value increases with decreasing angle θ. The latter gives us the intensity of the *information coupling* of the subjects in mutual communication.

This combined information can produce the emergence of the subjects participating in communication. And it happens in two cases: when related perceptions are grouped, $5 \cdot 2 + 3 \cdot 1 < (5 + 3) \cdot (2 + 1)$, or contradictory ones are decomposed, $5 \cdot 2 + (-3) \cdot (-1) > (5 - 3) \cdot (2 - 1)$. United allies are stronger, and defeated enemies are weaker.

2.2.3 Characteristic Equation

Let $\hat{A}$ be the operation of changing the load capacity of a freight vehicle λ times, and the state of the load is $|x\rangle$. Then the eigenequation of the work on the vehicle is $\hat{A}|x\rangle = \lambda|x\rangle$. Twice the applied operation of changing the load capacity on the vehicle is written as $\hat{A}^2|x\rangle = \lambda^2|x\rangle$. When the same n times is added to the load capacity increase, it will be $\hat{A}^n|x\rangle = \lambda^n|x\rangle$. The opposite operation to the first is defined by the inverse mapping $\hat{A}^{-1}$, such that $\hat{A}^{-1}|x\rangle = \lambda^{-1}|x\rangle$, so $\hat{A}^{-1}\hat{A} = \hat{I}$ means the statu quo, i.e. the unitary mapping. The operator $\hat{A}$ can represent a complicated engineering undertaking for the production of new machines, just as the eigenvalue equation of quantum mechanics $\hat{A}|x\rangle = \lambda|x\rangle$ can turn out to be a very demanding process of physical measurement. In addition, the quantum operator $\hat{A}$ is a hermitian, because it guarantees real values of the observable $\lambda \in \mathbb{R}$. In other words, Hermitian operators represent processes that are equal to themselves *dual*.

The Hermite (Charles Hermite, 1822 - 1901) operator $\hat{A}$ is self-adjoint if $\hat{A}^\dagger = \hat{A}$. In other words, it is self-adjoint if and only if $\langle \hat{A}x, y\rangle = \langle x, \hat{A}y\rangle$ for all $x, y \in X$. Recall that $\langle x| = |x\rangle^\dagger$ is the conjugate transpose of the vector $|x\rangle$. For example, in matrix

representation:

$$|x\rangle = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}, \quad \langle x| = \begin{pmatrix} \xi_1^* & \xi_2^* & \xi_3^* \end{pmatrix},$$

so it is:

$$\langle Ax, y\rangle = \begin{pmatrix} \xi_1^* & \xi_2^* & \xi_3^* \end{pmatrix} \begin{pmatrix} 3 & 2-i & -3i \\ 2+i & 0 & 1-i \\ 3i & 1+i & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} = \langle x, Ay\rangle,$$

because *adjoint*, i.e., the conjugate transposed matrix $A^\dagger = (A^*)^\top = A$, is equal to the original one. By conjugation, $(2-i)^* = 2+i$ and $i^* = -i$, and by transposing the matrix, types become columns. Applied to the same vector, it will be:

$$\lambda\|x\|^2 = \lambda\langle x, x\rangle = \langle \lambda x, x\rangle = \langle Ax, x\rangle = \langle x, Ax\rangle = \lambda^*\|x\|^2,$$

and hence $\lambda^* = \lambda$, which means $\lambda \in \mathbb{R}$.

Example 43. *Let us show that the eigenvalues of the given matrix A are $\lambda_1 = 6$, $\lambda_2 = -2$, and $\lambda_3 = -1$ in order. Then, the corresponding eigenvectors are:*

$$v_1 = \begin{pmatrix} 1-21i \\ 6-9i \\ 13 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1+3i \\ -2-i \\ 5 \end{pmatrix}, \quad v_3 = \begin{pmatrix} -1 \\ 1+2i \\ 1 \end{pmatrix}.$$

Here $Av = \lambda v$ and $i^2 = -1$.

Proof. First we look for the eigenvalues of λ. From $Av = \lambda v$, it follows that $(A - \lambda I)v = 0$, and hence the determinant is $\det(A - \lambda I) = 0$, so we have:

$$\begin{vmatrix} 3-\lambda & 2-i & -3i \\ 2+i & -\lambda & 1-i \\ 3i & 1+i & -\lambda \end{vmatrix} = 0,$$

$$(\lambda-6)(\lambda+2)(\lambda+1) = 0.$$

Hence $\lambda_1 = 6$, $\lambda_2 = -2$, and $\lambda_3 = -1$. Next, we solve the system of linear equations $Av = \lambda v$ three times, using the found eigenvalues, and for each finding the corresponding eigenvector. We get:

$$v_1 = \begin{pmatrix} 1-21i \\ 6-9i \\ 13 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1+3i \\ -2-i \\ 5 \end{pmatrix}, \quad v_3 = \begin{pmatrix} -1 \\ 1+2i \\ 1 \end{pmatrix}.$$

It is easier to check by direct multiplication that $Av_1 = 6v_1$, $Av_2 = -2v_2$ and $Av_3 = -v_3$, than to actually calculate. □

When we multiply (divide) the eigenvalue equation $Av = \lambda v$ by an arbitrary number α, we get again the eigenvalue equation $Au = \lambda u$, where $u = \alpha v$. This means that we can multiply the eigenvector by a (complex) number and still have an

eigenvector belonging to the same eigenvalue and again to the same matrix (operator). This justifies the custom in quantum physics of representing eigenvectors as normalized (divided by intensity).

A linear operator is *normal* if it commutes with its adjoint. Thus, when $AA^\dagger = A^\dagger A$, then A is normal. Operators commute with themselves, and, in particular, every Hermitian (self-adjoint) operator is normal, because then $A = A^\dagger$. But the converse is not true, as can be seen from:

$$NN^\dagger = \begin{pmatrix} 1 & 3 \\ -3 & 1 \end{pmatrix}\begin{pmatrix} 1 & -3 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix} = \begin{pmatrix} 1 & -3 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} 1 & 3 \\ -3 & 1 \end{pmatrix} = N^\dagger N.$$

Although it commutes with the adjoint, this matrix is not Hermitian: $N \ne N^\dagger$. However, $NN^\dagger$ is Hermitian, so it is always $\|Nx\| = \|N^\dagger x\|$, for every $x \in X$.

Theorem 44. *Normal and adjoint operators have the same eigenvectors.*

Proof. If $\hat{N}$ is normal, then $\hat{N} - \lambda\hat{I}$ is normal, so:

$$0 = \|(\hat{N} - \lambda\hat{I})v\| = \|(\hat{N} - \lambda\hat{I})^8 v\| = \|(\hat{N}^\dagger - \lambda^*\hat{I})v\|$$

and v is an eigenvector of both operators N and $N^\dagger$. The set of eigenvalues of an adjoint operator does not change under conjugation, so normal and adjoint operators have the same eigenvectors. □

Theorem 45. *When an operator is normal, the eigenvectors corresponding to its different eigenvalues are orthogonal.*

Proof. If α and β are the eigenvalues of the normal operator $\hat{N}$ of eigenvectors u and v respectively, then due to $\hat{N}^\dagger v = \beta^* v$, we have:

$$(\alpha - \beta)\langle u, v\rangle = \langle \alpha u, v\rangle - \langle u, \beta^* v\rangle = \langle \hat{N}u, v\rangle - \langle u, \hat{N}^\dagger v\rangle = 0,$$

then from $\alpha \ne \beta$ it follows that $u \perp v$. □

Because all self-adjoint (Hermitian) operators are normal, this statement also holds for them: the eigenvectors of the different eigenvalues of Hermitian operators are orthogonal. For example, the eigenvectors of the previous example (43) are orthogonal: $\langle v_1, v_2\rangle = 0$, $\langle v_1, v_3\rangle = 0$, and $\langle v_2, v_3\rangle = 0$. Hence, in quantum physics, we can always assign different eigenvalues to probabilities of finding different, independent *observables*.

The next theorem gives us another useful way to recognize these important, normal operators. This will be if both the operator and its adjoint change states to new ones of equal norm.

Theorem 46. *A linear operator $\hat{A}$ is normal if and only if $\|\hat{A}x\| = \|\hat{A}^\dagger x\|$ for all $x \in X$.*

Proof. It follows from the linearity of the operator $\hat{A}$ and the series of equivalences. If $\hat{A}$ is normal:

$$\hat{A}\hat{A}^\dagger - \hat{A}^\dagger\hat{A} = 0,$$

$$\langle(\hat{A}\hat{A}^\dagger - \hat{A}^\dagger\hat{A})x, x\rangle = 0,$$

$$\langle\hat{A}\hat{A}^\dagger x, x\rangle - \langle\hat{A}^\dagger\hat{A}x, x\rangle,$$

$$\|\hat{A}x\|^2 = \|\hat{A}^\dagger x\|^2,$$

for all $x \in X$. □

Considering Kepler's second law and, in general, central forces by which the motion of a charge sweeps out equal areas in equal times, we mentioned commutators (example 15). We connect commutators with the uncertainty relation $|[x,y]| \ge \varepsilon$, and this with the smallest and constant communication ε, or action between subjects x and y. In other words, when the action $\varepsilon > 0$, then we have some sustainable force that binds these subjects, and conversely, if $\varepsilon = 0$, the two $x \ne y$ are unconnected, or independent subjects. When processes commute, $xy = yx$, which is also true for states, there is no force or exchange of information between them. This is always the case of orthogonality $\langle x, y\rangle = 0$.

We have seen that the universe of physically real phenomena is one for which the law of conservation applies and, therefore, which builds its ever longer past so as to supplement the current present. The present is rarefied and moves away into its ever greater certainty, as if it were being consumed. Now, with knowledge from the inherent equations, values, and vectors, we can add another such "as." The forces of uncertainty that participate in the exchange of information between connected entities seem to consume the substance of the present. This is so formally correct that we can omit the adverb "as."

In the title 1.3.2 Vitalnost, the matrix A was an example of a simulation of the emergence of "something" from "nothing." It also cancels out a non-zero vector:

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

so the question arises: do such matrices exist in every space? The following paragraph gives a negative answer.

Theorem 47. *For a complex vector space X with scalar multiplication and a linear operator $\hat{A}$, the equality $\langle\hat{A}z, z\rangle = 0$ holds if and only if $\hat{A} = 0$.*

Proof. We decompose the scalar product $4\langle\hat{A}x, y\rangle$ into four sums:

$$\langle\hat{A}(x+y), x+y\rangle - \langle\hat{A}(x-y), x-y\rangle + i\langle\hat{A}(x+iy), x+iy\rangle - i\langle\hat{A}(x-iy), x-iy\rangle$$

which is true for all $x, y \in X$. Each of these is of the form $\langle\hat{A}z, z\rangle$, so from $\langle\hat{A}x, y\rangle$ it follows that each $\langle\hat{A}z, z\rangle = 0$, and hence $\hat{A} = 0$. □

In real spaces, this statement is also incorrect in the case of rotations by 90° around the origin. In complex space, the operator always rotates the vector radically. I once understood this remark as bypassing reality, let's say now, as the exit of a real particle from reality into the imaginary in order to survive in that way.

However, what has entered reality remains there, simply because "real"is what lasts, except in rare situations where probabilities dominate (in the micro world). Those familiar with algebra do not dare to give its interpretations in experimental situations, and those familiar with experimental abstractions of mathematics easily pass unnoticed. I guess that is how some simple algebraic truths remained without interpretations of reality. For example, such is the following statement. It says: once real, always real.

Theorem 48. *In a complex vector space X, with a dot product, a linear operator $\hat{A}$ is self-adjoint (Hermitian) if and only if $\langle \hat{A}x, x\rangle \in \mathbb{R}$ for every $x \in X$.*

Proof. If $x \in X$, then:

$$\langle \hat{A}x, x\rangle - \langle \hat{A}x, x\rangle^* = \langle \hat{A}x, x\rangle - \langle x, \hat{A}x\rangle = \langle \hat{A}x, x\rangle - \langle \hat{A}^\dagger x, x\rangle = \langle (\hat{A} - \hat{A}^\dagger)x, x\rangle.$$

If for each $\langle \hat{A}x, x\rangle \in \mathbb{R}$, then the origin of the equality is zero and $\langle (\hat{A} - \hat{A}^\dagger)x, x\rangle = 0$ for each vector $x \in X$. Hence, and by the previous statement, $\hat{A} - \hat{A}^\dagger = 0$. So, $\hat{A}$ is self-adjoint.

Conversely, if $\hat{A}$ is self-adjoint, then the last of the above equalities is zero, and so is the first, which means $\langle \hat{A}x, x\rangle \in \mathbb{R}$ for every $x \in X$. That is all that needed to be proved. □

The *scalar product* of the vector $\langle \hat{A}x, y\rangle$ is the information of the perception of what x will be after the process $\hat{A}$ with what y is, that is, it is the conjunction of the state x with y before that process. The scalar product is also a *functional*. Everything that is really active shows itself real in that interaction. Thus, the non-existent interactions of the present with the state after the process are a sign that that process did not exist. That is in the content of the next paragraph.

Theorem 49. *If $\hat{A} = \hat{A}^\dagger$ and $\langle \hat{A}x, x\rangle = 0$ for all x, then $\hat{A} = 0$.*

Proof. Let us start from the proof of Theorem 47, for all $x, y \in X$ and from:

$$\langle \hat{A}x, y\rangle = \frac{1}{4}[\langle \hat{A}(x+y), x+y\rangle - \langle \hat{A}(x-y), x-y\rangle]$$

which is now true, because the operator is self-adjoint. Hence $\hat{A} = 0$. □

The scalar product of the vectors $\langle a, b\rangle$ is interpreted as the internal *coupling* of the states a and b. This is the information of perception, which in its simple form is $\|a\|\|b\|\cos\phi$. The *pseudo scalar* product of the vectors $[a, b] = \|a\|\|b\|\sin\phi$ is interpreted as the external coupling, the trajectory of the charge state $a \to b$ with the center of constant force O, maybe zero force, but in the case of a trajectory in the form of a conic that decreases with the square of the distance. Considering that

this point O is anywhere, the external coupling is relative, so the internal coupling is also relative.

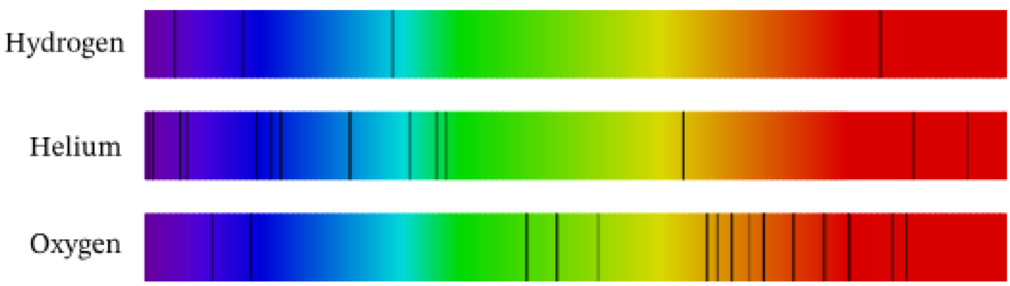
We write this in a concise way:

$$\langle a, b\rangle + i[a, b] = \|a\|\|b\|(\cos\phi + i\sin\phi) = \|a\|\|b\|e^{i\phi}. \tag{2.8}$$

If we consider this "angle" ϕ as a variable quantity, it will be that the internal and external coupling alternate, oscillating around the product of the intensities $\|a\|\|b\|$. This is enough for their "reality", a physical property of reality that in this way reveals its relativity and layering.

2.2.4 Spectrum

It is known how the composition of a material is determined based on the characteristics that appear in the spectrum of light coming from it (Spectra).

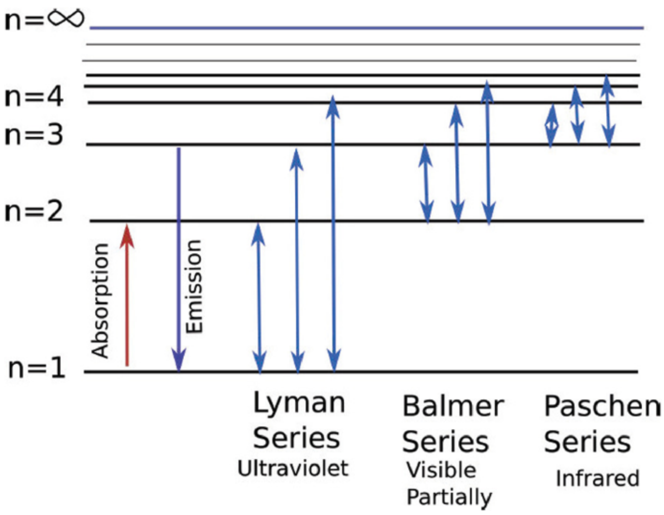


The emission spectrum of atomic hydrogen is in the differences of eigenvalues:

$$\frac{1}{\lambda} = R\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right),$$

where $R = 10\,973\,731.6$ m^{-1} is the Rydberg constant (see 110th example). The Lyman series is in the ultraviolet range. The Balmer series is in the visible range of the solar spectrum. The Paschen series is in the infrared range. According to Niels Bohr, the n-th eigenvalue of the Hermitian hydrogen operator $\hat{A}$ is $\lambda_n = -Rhc/n^2$, where $h = 6.62607015 \times 10^{-34}$ Js is Planck's constant, and $c = 299\ 792\ 458$ m/s is the speed of light. The spectra we see are the differences of such eigenvalues.

We used the numbers of energy levels $n = 1, 2, 3, \dots$ to solve the quantum harmonic oscillator by analyzing the Hamiltonian (2.2.1). Let us now look at these problems a little more broadly. First of all, let us recall that *adjoint* $\hat{A}^\dagger$ of the operator $\hat{A}$ means its conjugation and transposition, and that $\hat{A}^\dagger = \hat{A}$ characterizes a self-adjoint, i.e. Hermitian operator, which is almost exclusively used in quantum mechanics. When the vector space X is real, if the set of scalars $\Phi = \mathbb{R}$, then self-adjoint operators are *symmetric*.



The matrix representation A of the linear operator $\hat{A}$ depends on the choice of the basis of the vector space, hence on the choice of the coordinate system, among which:

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \end{pmatrix}, \quad \dots$$

and which we call the *standard basis*. It is common in quantum mechanics, but orthonormal bases are also used that can be obtained by *Gram-Schmidt procedure* orthogonalization. We often denote orthonormal basis vectors by Dirac kets $|n\rangle$, which also includes the standard basis.

The eigenvectors corresponding to different eigenvalues will be orthogonal if the matrix is symmetric. They are part of the *eigenspectrum*. We have already proved the theorem that interprets them (Theorem 45) for normal operators. They are of a broader kind that includes self-adjoint operators (Hermitian), or symmetric in the case of real spaces. The spectral theorem has several forms, all of which basically give the conditions under which a linear operator or matrix diagonalizes. In short, for a square matrix A of type $n \times n$ it is possible to find a basis in which it becomes diagonal if and only if A has n linearly independent eigenvectors.

These n linearly independent eigenvectors $v_1, v_2, \dots, v_n$ do not have to be orthogonal. This means that the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ do not have to be different ($Av_k = \lambda_k v_k$, where $k = 1, 2, \dots, n$). Therefore, the matrix may or may not be normal, and therefore neither Hermitian nor symmetric, in order to be reduced to a diagonal matrix. Such a square matrix A is similar to a diagonal matrix D with diagonal elements as eigenvalues of the given matrix. In this case, $A = PDP^{-1}$, and the matrix $P = P[v_1, v_2, \dots, v_n]$ is invertible and consists of columns of eigenvectors.

Example 50. *Let's diagonalize the matrix:*

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 1 \\ -1 & 0 & 1 \end{pmatrix},$$

if possible.

Solution. We are looking for the eigenvalues of the matrix:

$$\det(A-\lambda I)=0,$$

$$\begin{vmatrix} 2-\lambda & 0 & 0 \\ 1 & 2-\lambda & 1 \\ -1 & 0 & 1-\lambda \end{vmatrix} = 0,$$

$$(1-\lambda)(2-\lambda)^2 = 0.$$

The eigenvalues are $\lambda_1=1$, $\lambda_2=2$ and $\lambda_3=2$. The matrix is *degenerate*, its eigenvalues are not distinct and their associated eigenvectors will not be orthogonal.

We find three linearly independent eigenvectors of the matrix:

$$(A-\lambda_k I)v_k = 0, \quad k=1,2,3,$$

$$v_1 = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad v_3 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix},$$

so that $Av_k = \lambda_k v_k$. Check!

Therefore, there exists an invertible auxiliary matrix $P[v_1, v_2, v_3]$ whose columns are these eigenvectors and a diagonal matrix D, such that:

$$AP = PD,$$

$$\begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 1 \\ -1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 0 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix},$$

which is easy to check by direct multiplication. □

You will find many such examples in linear algebra classes. Some matrices with real entries that cannot be diagonalized over the set of real numbers $\mathbb{R}$ can be diagonalized over the complex numbers $\mathbb{C}$. For example:

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}\frac{1}{2}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}.$$

But there are matrices that cannot be diagonalized at all. For example:

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Here the eigenvalue $\det(A-\lambda I)v = 0$, i.e. the *characteristic* equation, reduces to $\lambda^2 = 0$. Its only eigenvalue is double $\lambda = 0$ and the eigenspace is one-dimensional, insufficient for a matrix basis of dimension two.

Matrices that have all repeated eigenvalues cannot be diagonalized, unless they are already diagonal. In such matrices we also find *eigen generalized* vectors. The eigenvector v of the matrix A belongs to the eigenvalue λ such that:

$$(A - \lambda I)v = 0.$$

However, a generalized eigenvector is such a w on which $A - \lambda I$ is nil-potent, with the same eigenvalue, but for some $n \in \mathbb{N}$:

$$(A - \lambda I)^n w = 0, \quad (A - \lambda I)^{n-1} w = v.$$

So, this is true for the generalized w and some eigenvector v, with the same eigenvalue λ of the matrix A.

Example 51. *Show that a matrix is degenerate*[2]*:*

$$A = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$$

and find the ordinary and generalized eigenvector.

Solution. The only eigenvalue, the solutions of the equation $\det(A - \lambda I) = 0$, are the double $\lambda_1 = \lambda_2 = 2$. Its one eigenvector is $v = (1, 0)$, and the other $w = (0, 1)$ can be taken to complete the basis. However, $(A - 2I)w = v$ and $(A - 2I)^2 w = 0$, so the second, $w = (0, 1)$, is a generalized eigenvector. □

Finally, note that there are complex symmetric matrices that cannot be diagonalized. For example, the non-Hermitian matrix $A^\dagger = (A^*)^\top \ne A$:

$$A = \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}$$

is symmetric $A^\top = A$. The eigenvalues, the solutions of the characteristic equation $\det(A - \lambda I) = 0$, are the conjugate complex numbers $\lambda_{1,2} = 1 \pm i$, but then the solutions $Av = \lambda v$ give zero vectors.

Spectral theorem of a complex vector space X with a linear operator $\hat{A}$ can be summarized in three equivalent statements:

1. $\hat{A}$ is normal;
2. X has an orthonormal basis consisting of the eigenvectors of $\hat{A}$;
3. $\hat{A}$ has a diagonal matrix with respect to some orthonormal basis X.

For a proof, further explanation, and applications of this theorem, open the attached link.

To make the spectral theorem easier to understand, let's note that the operator rotates, stretches, and maps states. By diagonalizing it, we express it as a set of

[2] degenerate matrix – has repeated eigenvalues

simple operators that do nothing more than stretch, or contract, the state v with the factors λ of the characteristic equations $\hat{A}v = \lambda v$. And this theorem tells us under what conditions a complete basis can be found for this vector space composed exclusively of eigenvectors:

$$X = X_1 + X_2 + ... + X_n, \quad X_k = \{v \in X \ : \ \hat{A}v = \lambda_k v\}.$$

What is sought is a coordinate system that, when rotated, makes the operator look very simple and can be expressed as a sum of multiplication operators that only act on the corresponding subspace of eigenvectors:

$$\hat{A} = \lambda_1 \hat{P}_1 + \lambda_2 \hat{P}_2 + ... + \lambda_n \hat{P}_n.$$

These operators greatly simplify the calculation by reducing it to identities and scalar multiplication. For a deeper understanding, add to all this the reality required by observables (physically measurable quantities) that are mutually distinct existing entities.

According to my information theory, I remind you, these different entities are unique couplings of interactions (2.8), of those who are subjects of communication, $\langle \mathbf{a}, \mathbf{b} \rangle = ab\cos\phi$ and of them in relation to the environment $[\mathbf{a}, \mathbf{b}] = ab\sin\phi$, where $a = \|\mathbf{a}\|$ and $b = \|\mathbf{b}\|$ are the capacities of the state of perception of the subjects. Secondly, what does not exist does not exist, which is to say that we divide the "existing"into "reality"and "fiction", otherwise two very complex and layered classes of entities. The first class includes all those that we can somehow perceive. They are connected in a long chain and therefore sustainable, permanent, or true, and the second class is the rest. There are unclear differences between some of these and we are working on that further.

2.2.5 Dual spaces

The *dual* of a vector space X over a field Φ is the space of linear functionals $f : X \to \Phi$. We denote it by $X^\dagger$, and it is also a vector space. A well-known example of a functional is $f_k(x) = \xi_k^*$, where $k = 1, 2, ..., n$, the projection of the vector $x = (\xi_1, \xi_2, ..., \xi_n)$ onto the k-th coordinate axis. We write this conjugate projection for short as $(x)_k = \xi_k^*$. It is clear that it is a linear mapping:

$$(ax + by)_k = [a(\xi_1, \xi_2, ..., \xi_n) + b(\eta_1, \eta_2, ..., \eta_n)]_k =$$

$$= [(a\xi_1, a\xi_2, ..., a\xi_n) + (b\eta_1, b\eta_2, ..., b\eta_n)]_k$$
$$= [(a\xi_1 + b\eta_1, a\xi_2 + b\eta_2, ..., a\xi_n + b\eta_n)]_k$$
$$= a\xi_k^* + b\eta_k^*,$$

so $(ax + by)_k = a(x)_k + b(y)_k$. Here the vector $y = (\eta_1, \eta_2, ..., \eta_n) \in X$, and it is assumed that the space is complex ($\Phi = \mathbb{C}$), so we do not omit conjugation.

The same applies to each of the projections. With basis vectors $e_1, e_2, ..., e_n \in X$, we can write the vector of the dual space as $x^\dagger = (x)_1 e_1 + (x)_2 e_2 + ... + (x)_n e_n$. It is

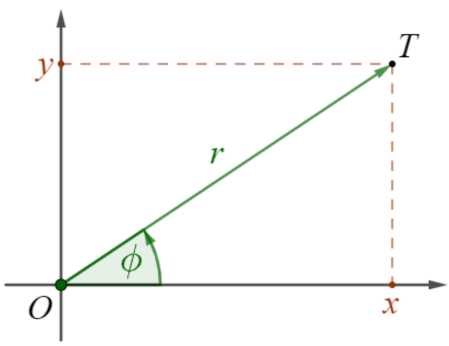
easy to see that the Dirac bra vector $\langle x|$ is actually dual to the ket vector $|x\rangle$ and that their product is the scalar product of the vector with itself, that is, the norm of that vector, $\langle x, x\rangle = x^\dagger \cdot x = \|x\|^2$. This shows that the basis vectors behave like dual vectors, that they are covariant duals of *contravariant* coordinates , and that their scalar product is invariant.

We interpret the space X as what is observed and the dual space $X^\dagger$ as what is observed, or vice versa; it doesn't matter because the dual of the dual space is the initial space ($X^{\dagger\dagger} = X$). One is the object, the object of observation, and the other is the subject with its observation. This interpretation sheds new light on self-adjunction and Hermitian operators. For example, what we observe is not what it is at that moment, because light takes some time to reach us.

Therefore, the immediate information of perception, the functional $\langle x, y\rangle$, we say the information coupling between the state x of the subject and the state y of the object, can be a complex number. Not even the past of the subject, for example, will be in the domain of real values for its present. But if the operator $\hat{A}$ is Hermitian and is a process that brings the past state of space into its present, then this coupling is real. Above is the proof (Theorem 48) that $\langle Ax, x\rangle \in \mathbb{R}$ for every $x \in X$ if and only if the operator $\hat{A}$ is self-adjoint.

Example 52. *Let's find the transformations of Cartesian rectangular coordinates (Oxy) into polar coordinates ($Or\phi$) and vice versa.*

Solution. In the figure on the left we see $x = r\cos\phi$ and $y = r\sin\phi$. Conversely, $r = \sqrt{x^2+y^2}$ and $\phi = \operatorname{arctg}(y/x)$. These are just the notations of the point T in different languages. □



The same matrix transformations are:

$$\begin{pmatrix} r \\ \phi \end{pmatrix} \to \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r\cos\phi \\ r\sin\phi \end{pmatrix}$$

and the inverse transformations:

$$\begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} r \\ \phi \end{pmatrix} = \begin{pmatrix} \sqrt{x^2+y^2} \\ \operatorname{arctg}\frac{y}{x} \end{pmatrix}.$$

I will use this example to list the tensor transformations that should then be connected to the previous one. We transform the operators of partial derivatives with:

$$\partial_x = \frac{\partial r}{\partial x}\partial_r + \frac{\partial \phi}{\partial x}\partial_\phi = \frac{x}{\sqrt{x^2+y^2}}\partial_r - \frac{y}{x^2+y^2}\partial_\phi = \cos\phi\,\partial_r - \frac{\sin\phi}{r}\,\partial_\phi,$$

$$\partial_y = \frac{\partial r}{\partial y}\partial_r + \frac{\partial \phi}{\partial y}\partial_\phi = \frac{y}{\sqrt{x^2+y^2}}\partial_r + \frac{x}{x^2+y^2}\partial_\phi = \sin\phi\,\partial_r + \frac{\cos\phi}{r}\,\partial_\phi.$$

In general, quantities that transform in this way:

$$\begin{cases} A_x = A_r \frac{\partial r}{\partial x} + A_\phi \frac{\partial \phi}{\partial x} \\ A_y = A_r \frac{\partial r}{\partial y} + A_\phi \frac{\partial \phi}{\partial y} \end{cases}$$

are *covariant tensors* of the first rank, or vectors. They have subscripts. The coordinates themselves, however, are transformed in the following way:

$$dx = \frac{\partial x}{\partial r} dr + \frac{\partial x}{\partial \phi} d\phi = \cos\phi\, dr - r\sin\phi\, d\phi,$$

$$dy = \frac{\partial y}{\partial r} dr + \frac{\partial y}{\partial \phi} d\phi = \sin\phi\, dr + r\cos\phi\, d\phi.$$

Generally, quantities that are transformed as coordinates:

$$\begin{cases} A^x = A^r \frac{\partial x}{\partial r} + A^\phi \frac{\partial x}{\partial \phi} \\ A^y = A^r \frac{\partial y}{\partial r} + A^\phi \frac{\partial y}{\partial \phi} \end{cases}$$

become contravariant vectors, or tensors of the first rank. We write them with the upper indices.

An example of a covariant tensor of the first rank is $-i\hbar(\partial_x, \partial_y, \partial_z)$. This is the *momentum operator*. A common example is the *metric tensor*, otherwise twice covariant, which we find by calculating the infinitesimal interval, the Pythagorean in other coordinates:

$$(d\ell)^2 = (dx)^2 + (dy)^2 = (\cos\phi\, dr - r\sin\phi\, d\phi)^2 + (\sin\phi\, dr + r\cos\phi\, d\phi)^2.$$

After squaring, adding, and arranging, we get

$$d\ell^2 = dr^2 + r^2 d\phi^2.$$

Following the meaning, we will not be confused with squaring with superscripts. Continuing, we generalize the above to:

$$ds^2 = g_{ij} dx^i dx^j,$$

where, according to Einstein's convention, addition by repeated subscripts and superscripts is implied. Namely, from the general coordinate transformations:

$$x_\mu = x_\mu(\bar{x}_1, \bar{x}_2, ..., \bar{x}_n),$$

$$dx_\mu dx_\nu = \frac{\partial x_\mu}{\partial \bar{x}_i} \frac{\partial x_\nu}{\partial \bar{x}_j} d\bar{x}_i d\bar{x}_j,$$

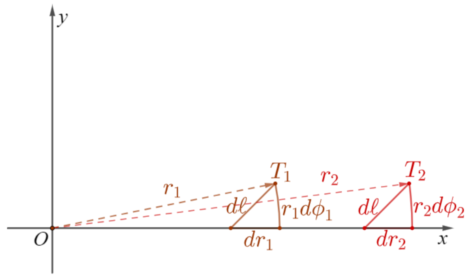
it follows:

$$g_{ij} = \frac{\partial x_\mu}{\partial \bar{x}_i} \frac{\partial x_\nu}{\partial \bar{x}_j},$$

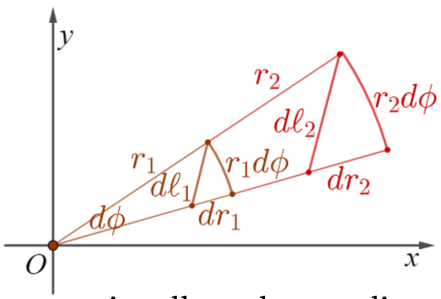
and hence the g_{ij} tensor is twice covariant.

This interval (ds) is an *invariant* of the coordinate transformation, but because it is obtained by multiplying twice the covariant metric tensor (g_{ij}) by twice the contravariant coordinates ($dx^i dx^j$). When, as in polar coordinates, no mixed products appear in the expression for the interval (all $g_{ij} = 0$ if $i \neq j$), then the coordinates are orthogonal.

The figure on the right shows what the interval $d\ell$ represents in the case of polar coordinates. The points T_1 and T_2 are at distances r_1 and r_2 from the origin O. When the infinitesimal intervals dr_1 and dr_2, also $r_1 d\phi_1$ and $r_2 d\phi_2$, are equal so that the diagonals $d\ell$ are equal to each other, then this interval is invariant. We find this when we assume that the infinitesimals dx and dy of rectangular coordinates are invariant if these coordinates increase and given that $dx^2 + dy^2 = dr^2 + r^2 d\phi^2$.



The picture is different if we consider that $d\ell$ grows proportionally to r, which is usually assumed in calculus. This can be seen in the figure on the left, where dr grows with r, but the angle $d\phi$ does not change with the change in ϕ. Then, again, from the equality



$$dr^2 + r^2 d\phi^2 = dx^2 + dy^2,$$

it follows that the infinitesimals dx and dy grow proportionally to the coordinates x and y, so the surface $dxdy$, or $rdrd\phi$, also grows. This is usually determined using the Jacobian, the determinant which in the 2-dim case represents the alignment of surfaces created by stretching the coordinates.

We achieve the same using the commutator:

$$[\partial x, \partial y] = \partial_r x \partial_\phi y - \partial_\phi x \partial_r y = \cos\phi \cdot r\cos\phi - r\sin\phi \cdot \sin\phi = r.$$

The Jacobian of polar coordinates is $J(r, \phi) = r$. In the case of a 3-dimensional coordinate system, $(x_1, x_2, x_3) \to (\bar{x}_1, \bar{x}_2, \bar{x}_3)$, the Jacobian is the volume:

$$J(\bar{x}_1, \bar{x}_2, \bar{x}_3) = \begin{vmatrix} \frac{\partial x_1}{\partial \bar{x}_1} & \frac{\partial x_1}{\partial \bar{x}_2} & \frac{\partial x_1}{\partial \bar{x}_3} \\ \frac{\partial x_2}{\partial \bar{x}_1} & \frac{\partial x_2}{\partial \bar{x}_2} & \frac{\partial x_2}{\partial \bar{x}_3} \\ \frac{\partial x_3}{\partial \bar{x}_1} & \frac{\partial x_3}{\partial \bar{x}_2} & \frac{\partial x_3}{\partial \bar{x}_3} \end{vmatrix}.$$

For example, the Jacobian of Cartesian rectangular *coordinates* is $J(x, y, z) = 1$, and of spherical coordinates $J(\rho, \varphi, \theta) = \rho^2 \sin\varphi$, because for infinitesimal volumes in these two systems the equality $dxdydz = \rho^2 \sin\varphi d\rho d\varphi d\theta$ holds.

We use these alignments due to coordinate stretching, for example, when calculating the double integral for a 2-dimensional space:

$$\iint_A dx_1 dx_2 = \iint_B |J(\bar{x}_1, \bar{x}_2)| \, d\bar{x}_1 d\bar{x}_2$$

and analogously with some other number of dimensions.

Tensors of rank zero, upper invariants, whether they are formed by of tensors of rank one, two, or higher:

$$A_i B^i = A_1 B^1 + A_2 B^2 + ... + A_n B^n, \quad A_{ij} B^{ij} = A_{11} B^{11} + A_{12} B^{12} + ... + A_{mn} B^{mn},$$

are types of mappings of tensor quantities (vectors, matrices, and more complex ones) into scalars. Therefore, they are *functionals*, but not necessarily linear.

2.2.6 Ricci tensor

Next to the metric tensor g_{ij}, the immediately important Ricci tensor R_{ij} (1887-1896) is also doubly covariant, not so simple, but good for understanding tensor calculus in general and Einstein's theory of general relativity in particular. One by one, I will explain the contravariant metric tensor, then the Christoffel symbols, then the Riemann tensor, and finally the Ricci tensor. That which concerns the mathematics itself, and then the application of it in the form of Einstein's equations.

If the metric tensor is viewed as a matrix, then $g = g(x, \bar{x}) = (g_{ij})$ and $g^{-1} = (g^{ij})$ are inverse matrices, $gg^{-1} = g^{-1}g = I$. Their coefficients are:

$$g_{ij} = \frac{\partial x_\mu}{\partial \bar{x}_i}\frac{\partial x_\nu}{\partial \bar{x}_j}, \quad g^{ij} = \frac{\partial \bar{x}_\mu}{\partial x_i}\frac{\partial \bar{x}_\nu}{\partial x_j}.$$

This is often taken as the definition of a contravariant metric tensor, that $g_{ij}g^{jk} = \delta_i^k$, where $\delta_i^k = 1$ only if $i = k$, and zero otherwise. The invariant obtained by contracting a co- and contravariant metric tensor is the identity matrix. By contracting the identity matrix, $\delta_i^i = n$, one obtains the order of the matrix.

Example 53. *Let us show that*

$$g_{\alpha\beta} = \frac{\partial x^i}{\partial y^\alpha}\frac{\partial x^j}{\partial y^\beta} g_{ij}$$

is the transformation rule of the metric tensor.

Proof. Moving to the new coordinates, $x \to y$, we find:

$$ds^2 = g_{ij}dx^i dx^j = g_{ij}\frac{\partial x^i}{\partial y^\alpha}dy^\alpha \frac{\partial x^j}{\partial y^\beta}dy^\beta =$$

$$= g_{ij}\frac{\partial x^i}{\partial y^\alpha}\frac{\partial x^j}{\partial y^\beta}dy^\alpha dy^\beta = g_{\alpha\beta}dy^\alpha dy^\beta.$$

Hence the sought equality. □

The components of the metric tensor are defined by scalar dot products between the basis vectors, $g_{ij} = \vec{e}_i \cdot \vec{e}_j$, so that the diagonal elements ($i = j$) describe the squares of the lengths of the basis vectors, and the off-diagonal ones ($i \ne j$) describe the overlap of the i-th basis vector with the j-th. In the case of orthogonality ($\vec{e}_i \perp \vec{e}_j$) of different ones, such a metric coefficient is zero ($g_{ij} = 0$), and when all coordinates are mutually orthogonal, then the matrix of the metric tensor is diagonal. For example, as in the case of polar coordinates:

$$(g_{\mu\nu}) = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}, \quad (g^{\mu\nu}) = \begin{pmatrix} 1 & 0 \\ 0 & 1/r^2 \end{pmatrix}$$

then it is easy to calculate the inverse, contravariant metric tensor.

Determinant of a matrix of order n is otherwise the product of n eigenvalues of the matrix and represents the n-dim "volume,"which is the surface in the case $n = 2$. When we compare it with the above-mentioned Jacobian, it will be $J = \sqrt{\det(g_{\mu\nu})}$.

In general *theory of relativity*, the space-time coordinates dx^1, dx^2, dx^3 and $dx^4 = d(ict)$, where $i^2 = -1$, and $c \approx 300\ 000$ km/s is the speed of light, the metric tensor represents gravitational potentials. The important thing here is g_{44}, which gives time dilation, because gravity pulls where time flows more slowly.

Christoffel symbols (1869) of differential geometry are coefficients that describe parallel transport in curvilinear coordinates. Those of the first kind are denoted by $[\mu\nu, \kappa] = \Gamma_{\mu\nu\kappa}$, and Christopher symbols of the second kind are $\begin{Bmatrix} \sigma \\ \mu\nu \end{Bmatrix} = \Gamma^\sigma_{\mu\nu}$. They are not tensors because they do not transform as co- or contra-variant quantities. Physically, Christopher's symbols represent some fictitious, spontaneous forces caused by a non-inertial reference frame.

Christoffel symbols can be interpreted geometrically as descriptions of changes in the basis vectors in a given coordinate system. These vectors change due to the curvilinear coordinate system or due to the geometry of the space itself that is curved, and Christoffel symbols describe both of these. Essentially they are derivatives of the basis vectors:

$$\Gamma^\sigma_{\mu\nu} = \frac{\partial \vec{e}_\mu}{\partial x^\nu} \cdot \vec{e}^\sigma$$

$$\Gamma^\sigma_{\mu\nu} = \vec{e}^\sigma \cdot \partial_\nu \vec{e}_\mu \implies \partial_\nu \vec{e}_\mu = \Gamma^\sigma_{\mu\nu} \vec{e}_\sigma$$

after multiplying by $\vec{e}_\sigma$ on both sides and using $e_\sigma e^\sigma = 1$. When we sum over σ, it becomes clear that the Christoffel symbols represent the components of the vector $\partial_\nu \vec{e}_\nu$. That is why they are called the connection coefficients (of the basis and derivative vectors). After a more extensive calculation, from this geometric definition, we obtain Christoffel symbols of the second kind expressed using the metric tensor:

$$\Gamma^\sigma_{\mu\nu} = \frac{1}{2} g^{\sigma\alpha} (\partial_\mu g_{\alpha\nu} + \partial_\nu g_{\mu\alpha} - \partial_\alpha g_{\mu\nu}).$$

They are symmetric in the lower two indices, $\Gamma^\sigma_{\mu\nu} = \Gamma^\sigma_{\nu\mu}$. You can find the calculation of this formula and something else about these symbols in my attachment by following the link above. Since $\Gamma^\sigma_{\mu\nu} = g^{\sigma\alpha} \Gamma_{\alpha\mu\nu}$, these are the Christoffel symbols of the first kind:

$$\Gamma_{\alpha\mu\nu} = \frac{1}{2} (\partial_\mu g_{\alpha\nu} + \partial_\nu g_{\mu\alpha} - \partial_\alpha g_{\mu\nu}).$$

In polar coordinates $x^1 = r$, $x^2 = \phi$ the Christoffel symbols are:

$$\Gamma^1_{22} = -r, \quad \Gamma^2_{12} = \Gamma^2_{21} = \frac{1}{r},$$

and all others are zero. Written in matrix form:

$$\Gamma^1_{ij} = \begin{pmatrix} 0 & 0 \\ 0 & -r \end{pmatrix}, \quad \Gamma^2_{ij} = \begin{pmatrix} 0 & \frac{1}{r} \\ \frac{1}{r} & 0 \end{pmatrix}.$$

These are well-known details of tensor calculus, so I only mention them.

Long before this discovery, Gauss (1827) found the Theorema Egregium when he defined the curvature of a surface $K = 1/(r_1 r_2)$ by means of tangential to the surface and mutually perpendicular circles oriented with radii r_1 and r_2. On a sphere, such are the greatest circles of radius $r_1 = r_2 = r$, so the *Gaussian curvature* of a sphere is a constant $1/r^2$. On a saddle surface, these are circles in opposite directions with negative curvature $-1/(r_1 r_2)$. On a cylinder, one of the two circles has an infinite radius, so the cylinder has zero curvature. It therefore develops into a plane without plastic deformation of its envelope, while this is not possible with a spherical or saddle surface.

Gauss' discovery was translated by his student Riemann (1854) into Riemann tensor to apply and express these curves in more detail:

$$R^\alpha_{\beta\gamma\delta} = \partial_\gamma \Gamma^\alpha_{\beta\delta} - \partial_\delta \Gamma^\alpha_{\beta\gamma} + \Gamma^\mu_{\beta\delta}\Gamma^\alpha_{\mu\gamma} - \Gamma^\mu_{\beta\gamma}\Gamma^\alpha_{\mu\delta}.$$

A purely covariant version of this curvature tensor is:

$$R_{\beta\gamma\delta\rho} = g_{\rho\zeta} R^\zeta_{\beta\gamma\delta}.$$

In one dimension $R_{1111} = 0$, but in four dimensions it has 256 components. However, there are symmetries that simplify their calculations:

$$R_{iklm} = -R_{ikml} = -Rkilm.$$

These reduce the number of components to 36, and with:

$$R_{iklm} = R_{lmik}$$

this number is reduced to 21. The last symmetry:

$$R_{iklm} + R_{ilmk} + R_{imkl} = 0$$

reduces the number of components to 20. In general, in n dimensions, the number of independent components will be $n^2(n^2 - 1)/12$.

The interesting thing about the Riemann tensor is that the contraction of one pair of indices gives the Ricci tensor:

$$R_{ij} = R^k_{ikj} = R^k_{kij} = R^k_{jki} = R_{ji},$$

$$R_{ij} = g^{k\mu} R_{\mu ikj}, \quad R^k_j = g^{ik} R_{ij}.$$

The contraction of the Ricci tensor gives the Ricci curvature scalar $R = R^k_k$. Since the Riemann tensor is identically zero in flat spacetime, the Ricci tensor and curvature scalar are also zero. However, the Riemann tensor always has at least one nonzero component in curved spacetime, so the Ricci tensor and scalar can be zero in curved spacetime as well. So if the Ricci tensor or scalar is nonzero, the spacetime is curved, but if it is zero, then we need to work out the full Riemann tensor to see if there is curvature or not.

If the Ricci scalar curvature R is positive, it means that the space (manifold) tends to curve like a sphere; if it is negative, it is more of a saddle shape. In 2-dimensional space, the Ricci tensor is more directly related to the scalar curvature because they are essentially the same; however. The Ricci tensor tells us how the volume changes relative to the Euclidean, in different directions at a point in the manifold. Both the Ricci tensor and the scalar are important parts of the equations of general relativity, Einstein's tensors:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R, \quad G_{\mu\nu} + \Lambda g_{\mu\nu} = \kappa T_{\mu\nu}, \quad \kappa = \frac{8\pi G}{c^4} \approx 2.1 \times 10^{-43}\ \mathrm{N}^{-1},$$

which describe the curvature of mass, energy ($E_{\mu\nu}$) and space-time ($G_{\mu\nu}$).

Of the listed properties, I emphasize: The Ricci tensor represents the difference between the volume of a given curved space and the corresponding volume of Euclidean space. The Ricci tensor measures volumes in units of geodesics (shortest paths) and their changes due to curvature. In general relativity, the Ricci tensor is a measure of volume changes due to gravitational tides.

2.3 Stochastic processes

We have seen that quantum processes (Hermitian) can walk through complex states, with some purely imaginary values, and yet always have real observables. We did not pay attention to their stochastic (random) nature, often their central place. We did not emphasize it by pursuing dual spaces, processes, and states, or we did connect them with the information of perception, but even then only in hints. Especially we did not notice this probabilistic nature of things in the theory of relativity when we almost overlooked that tensors are components of states and processes because they build vectors, matrices, and even more complex quantities.

The previous two titles (2.1 Vectors and 2.2 Processes) revolve around functionals, from scalar products to perceptual information to zero-rank tensors, and we will continue this trend by dealing with probabilities more intensively.

2.3.1 Conditional probability

Experiment is a phenomenon, random event, or experiment in probability theory where two or more outcomes are possible. An example is tossing a coin with the outcomes {tails, heads}, or rolling a die with the outcome "the number of dots on the top side when the die falls." The set of all logically possible outcomes of an experiment is denoted by Ω, and its elements, individual outcomes, by ω.

By defining events as elements of a set, $\omega \in \Omega$, we are free to use the familiar properties of *sets*. Thus, A implies B we write $A \subset B$ if and only if the realization of event A also implies the realization of event B. The complement of the set $A \subset \Omega$ is the set $A^c \subset \Omega$ whose trials are realized if and only if the trials of the set A are not realized. The intersection operation $A \cap B$ (of the product AB) is an event that

is realized iff[3] both A and B are realized. When $AB = \varnothing$, the events A and B are disjoint; they are mutually exclusive; they cannot be realized simultaneously. The union operation $A \cup B$ (the sum of $A + B$) is an event that occurs if at least one of the events A and B occurs. For Ω, we say that is a *sure event*, and for $\varnothing$, we say that is an *impossible event*.

Definition 54. *. A class of events $\mathcal{F}$ constitutes a σ-field (σ-algebra) when: 1. $\Omega \in \mathcal{F}$, 2. from $A \in \mathcal{F}$ follows $A^c \in \mathcal{F}$, 3. if $A_k \in \mathcal{F}$, k = 1, 2, ..., then $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$.*

Taking the complement gives $\varnothing \in \mathcal{F}$, and from $A_k \in \mathcal{F}$, $k = 1, 2, ...$, it follows $\bigcap_{k=1}^{\infty} A_k \in \mathcal{F}$. In probability theory, one considers some collection of events C that is complemented by the complement up to Ω and thus obtains a σ-field. Usually we work with the minimal σ-field that contains a given collection C, say that it is generated by the collection C, and it turns out that such a field always exists.

Theorem 55. *A minimal σ-field $\mathcal{F}(C)$ containing a non-empty collection C always exists.*

Proof. Let $S = \{\mathcal{S}_\lambda\}$ be the set of all σ-fields containing C. It is not an empty set, because all subsets of Ω form a single σ-field containing C. Also, the σ-field is the intersection of $\bigcap_\lambda \mathcal{S}_\lambda$. However, it is minimal, because for every λ' there exists $\mathcal{S}_{\lambda'}$, and yet $\bigcap_\lambda \mathcal{S}_\lambda \subset \mathcal{S}_{\lambda'}$. This completes the proof. □

Next, we consider only experiments in which the events from the σ-field $\mathcal{F}$ possess some statistical stability, which can be repeated arbitrarily many times n and in each repetition we register whether an event $A \in \mathcal{F}$ has been realized or not. Let $n(A)$ be the number of realizations of the event A in n repeated experiments. Then the *relative frequency* (statistical probability) of that event is the quotient $n(A)/n$. It is shown that as the number $n \to \infty$ increases, the relative frequency tends to cluster, $n(A)/n \to \Pr(A)$, around some fixed number $\Pr(A)$, which we call the probability of the event A. We denote probabilities with the letters $P, Q, ...$, or lowercase $p, q, ...$, and even with other symbols when the text makes it clear what is being referred to.

Definition 56. *$P(.)$ is a numerical function defined over the σ-field of events $\mathcal{F}$, which has the following properties:*

1. (non-negativity) for every $A \in \mathcal{F}$ $P(A) \ge 0$;

2. (normativity) $P(\Omega) = 1$;

3. (σ-additivity) for mutually disjoint events ($A_i A_j = \varnothing$ when $i \ne j$) from $A_k \in \mathcal{F}$ for all $k = 1, 2, ...$ it follows:

$$P\left(\sum_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k).$$

These are the probability axioms of Kolmogorov (1933). Although there are only three of them, they will serve to derive all the necessary consequences. I will list nine of them:

[3] iff – if and only if

i. $P(\varnothing) = 0$ follows from $\Omega + \varnothing = \Omega$;
ii. finite additivity:

$$P\left(\sum_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k)$$

follows from $\sum_{k=1}^{n} A_k = A_1 + ... + A_n + \varnothing + ...$;
iii. if $A \subseteq B$ then $P(A) \le P(B)$, it follows from $B = A + A^c B$;
iv. $\forall A \in \mathcal{F}$ is $0 \le P(A) \le 1$, it follows from $\varnothing \subseteq A \subseteq \Omega$;
v. $P(A^c) = 1 - P(A)$, it follows from $A + A^c = \Omega$;
vi. $P(A \cup B) = P(A) + P(B) - P(AB)$, from $A \cup B = A + A^c B$ and $B = AB + A^c B$;
vii. *covering lemma*:

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) \le \sum_{k=1}^{\infty} P(A_k),$$

follows from:

$$\bigcup_{k=1}^{\infty} A_k = A_1 + A_1^c A_2 + A_1^c A_2^c A_3 + ..., \quad A_1^c ... A_{k-1}^c A_k \subseteq A_k, \quad k = 2, 3, ...;$$

viii. continuity of the function P: if the sequence of events $A_1, A_2, ...$ is monotonically non-decreasing, i.e. $A_1 \subseteq A_2 \subseteq ...$, then

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{n\to\infty} P(A_n).$$

Proof. Let us write the union as a sum of disjoint events:

$$\bigcup_{k=1}^{\infty} A_k = A_1 + A_1^c A_2 + A_2^c A_3 + ...,$$

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = P\left(\sum_{k=1}^{\infty} A_{k-1}^c A_k\right) = \sum_{k=1}^{\infty} P(A_{k-1}^c A_k) =$$

$$= \lim_{n\to\infty} \sum_{k=1}^{n} P(A_{k-1}^c A_k) = \lim_{n\to\infty} P\left(\sum_{k=1}^{n} A_{k-1}^c A_k\right) = \lim_{n\to\infty} P(A_n), \quad (A_0 = \varnothing)$$

and that was to be proven. □

Moving on to the opposite events, it is also easy to prove: If the sequence of events $A_1, A_2, ...$ is monotonically non-increasing, i.e. $A_1 \supseteq A_2 \supseteq ...$, then

$$P\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{n\to\infty} P(A_n).$$

ix. Borel-Kantelli lemma (first): if the sequence

$$\sum_{k=1}^{\infty} P(A_k)$$

converges, then with probability 1 only finitely many events from the sequence $A_1, A_2, ...$ can be realized.

Example 57. *Three asteroids A, B and C independently orbit the solar system with chances of 5, 10 and 15 percent respectively of hitting Earth in the next millennium. What is the probability that at least one of them will hit us during that period?*

Solution. The probabilities of hits are $P(A) = 0.05$, $P(B) = 0.10$, and $P(C) = 0.15$. The probabilities of misses are $P(A^c) = 0.95$, $P(B^c) = 0.90$, and $P(C^c) = 0.85$, respectively. Therefore, the probability that at least one of them will hit us in that period is $1 - P(A^c B^c C^c) = 1 - 0.95 \cdot 0.90 \cdot 0.85 = 1 - 0.72675 = 0.2732$. That's about 27 percent. □

Example 58. *It is found that 42 out of 100 men and 64 out of 100 women wear glasses. The number of women in a certain room is four times the number of men and a person is chosen at random. Calculate the probability that:*

A. The person is without glasses.

B. A woman with glasses is chosen.

Solution. The first is like a mean value question:

$$P(A) = \frac{1}{5} \cdot \frac{58}{100} + \frac{4}{5} \cdot \frac{36}{100} = \frac{58 + 144}{500} = \frac{202}{500} = 0.404.$$

$$P(B) = \frac{4}{5} \cdot \frac{64}{100} = 0.512.$$

The second is a matter of the 4 : 5 participation ratio of the group with 64 percent of the glasses. □

It is said that we observe random events that possess some "statistical stability". Given the definition of *reality* as something that can:

1. be perceived through a series of intermediaries, or
2. can last, that is,
3. which is true,

then there is also *statistical stability*. Namely, the chain of perceptions is a process. It is a flow that speaks of the second determinant of "reality". Again, what is not true will not be shown in a physical experiment, so there is a third determinant.

On the other hand, what lasts leaves its mark, has a past that acts on the present, directing it and thereby limiting it. The more history there is in the perception of the present, the less information (the amount of options) there is in the current events themselves – the law of conservation of information dictates. That this is not a personal matter of long periods of cosmic development, we will see further with the help of *conditional probabilities*.

If we know or assume that an event A has occurred, then this may have an impact on the frequency of another event B. Let the relative frequencies of the occurrence of events A, B and AB be the numbers n_A, n_B and n_{AB} in n trials, respectively. Then the conditional relative frequency of the event B in trials that certainly lead to the occurrence of event A is:

$$\frac{n_{AB}}{n_A} = \frac{n_{AB}/n}{n_A/n}.$$

Accordingly, conditional probabilities should be defined as follows:

$$P(B|A) = \frac{P(AB)}{P(A)}, \quad P(A) > 0. \tag{2.9}$$

Theorem 59. *$P(\cdot|A)$, $P(A) > 0$ is the probability over $(\Omega, \mathcal{F})$ with $P(A|A) = 1$.*

Proof. The properties of probability are determined by definition 56. and the definition of $P(\cdot|A)$. The first axiom is obvious, $P(B|A) \ge 0$, because such is the quotient $P(AB)/P(A)$. The second is also obvious, $P(A|A) = P(AA)/P(A) = 1$. Based on the distribution of the intersection with respect to the union, we have:

$$\left(\sum_{k=1}^{\infty} B_k\right) A = \sum_{k=1}^{\infty} B_k A,$$

$$P\left(\sum_{k=1}^{\infty} B_k | A\right) = \frac{P\left(\sum_{k=1}^{\infty} B_k A\right)}{P(A)} = \sum_{k=1}^{\infty} \frac{B_k A}{A} = \sum_{k=1}^{\infty} B_k A.$$

This is the third axiom and the proof is complete. □

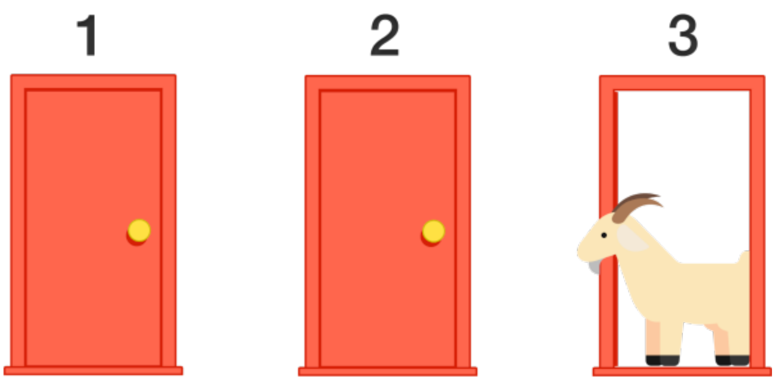
Example 60. *The probability that a product will work properly during the first t years of use is $\exp(-t/6)$. The product worked for the first three years. What is the probability that it will break down during the fourth year?*

Solution. Let A be the event that the product breaks down in the fourth year, and B that it does not break down during the first three years. We find:

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(A)}{P(B)} = \frac{\exp(-3/6) - \exp(-4/6)}{\exp(-3/6)} \approx 0.1535.$$

The chance of a breakdown exactly during the fourth year is about 15 percent. □

Monty Hall's Paradox is the next problem. It is a form of probability puzzle based on the American television game show "Let's Make a Deal"and named after its original host, Monty Hall.



You are in a quiz and are asked to choose one of three doors. Behind each door is either a car or a goat. You choose the door. The host, Monty Hall, chooses one

of the other doors, which he knows has a goat behind it, and opens it, showing you the goat (he deliberately does not choose the car). Monty then asks if you would like to change your door choice to the other remaining doors. Assuming you would rather have the car than the goat, do you choose the change or not?

The solution is that switching will allow you to win twice as often as sticking with your original choice, a result that seems counterintuitive to many. The Monty Hall problem famously embarrassed a number of mathematicians with PhDs when they attempted to "correct" Marilyn vos Savant's solution in a column in Parade magazine.

Example 61. *Why is it better to choose one of the two doors again, when Monty opens the third one with the goat behind it?*

Explanation. Simply put, because then you choose a car with a 50% chance, and before that the chance was 30%. By choosing randomly then, e.g. by flipping a coin, it may turn out to be the same original door, but part of the uncertainty thus belongs to the others as well. In longer series, by repeating the choice with a coin, this excess uncertainty corresponds exactly to the difference in probabilities 1/2 - 1/3, i.e. the increase in the chance of winning by choosing one of the two doors, after Monty opens the third one (behind which is the goat). □

2.3.2 Probability of hypotheses

Using the conditional probabilities (2.9), we define *product probability* of two events:

$$P(AB) = P(A)P(B|A) = P(B)P(A|B). \tag{2.10}$$

The last equality follows from the commutativity of the product $AB = BA$. We will return to this shortly, due to Theorem 67.

We say that disjoint events $A_1, A_2, ..., A_n \in \Omega$ form an *event decomposition* of Ω if

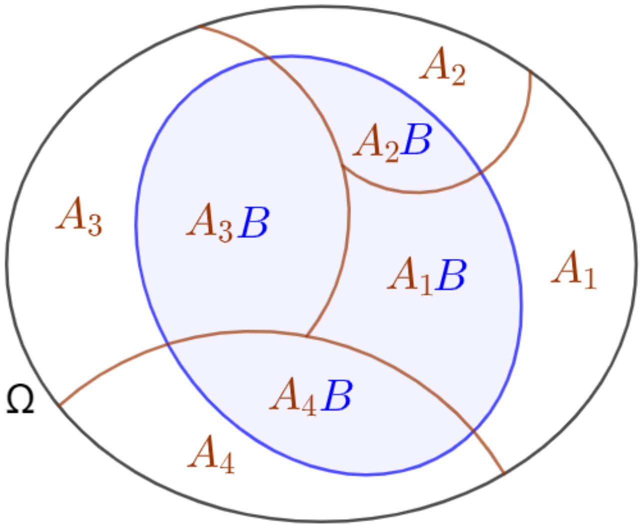
$$\sum_{k=1}^{n} A_k = \Omega,$$

i.e. if one and only one of the events $A_1, A_2, ..., A_n$ is always realized. In the following figure we see what happens when we intersect an arbitrary event $B \subset \Omega$, or multiply it by some of the "event decompositions" of the same Ω. We obtain disjoint events $A_k B \subset \Omega$.

The formula *complete probability* reads: if the events $A_1, A_2, ..., A_n$ form one decomposition of Ω then:

$$P(B) = \sum_{k=1}^{n} P(A_k)P(B|A_k). \tag{2.11}$$

The proof follows directly from the equality $B = \sum A_k B$, so $P(B) = \sum P(A_k B)$ and by extension $P(A_k B) = P(A_k)P(B|A_k)$.



Example 62. *Passengers from three cities pass through the checkpoint. They are three times more likely to come from the first than from the second and third cities, and the frequencies of those from those two are equal. It is estimated that 5 percent of those from the first city are smuggling something, while the number of smugglers from the second and third cities is 3 percent each. One of the passengers is stopped at random. Find the probability that he is a smuggler.*

Solution. If A_k is the event "the traveler is from k-th city", then $P(A_1) = 3P(A_2)$, $P(A_2) = P(A_3)$. Since $P(A_1) + P(A_2) + P(A_3) = 1$ we calculate $P(A_1) = 3/5$ and $P(A_2) = P(A_3) = 1/5$. If the event B is: "the traveler is a smuggler", we have conditional probabilities $P(B|A_1) = 0.05$ and $P(B|A_2) = P(B|A_3) = 0.03$. Then, based on the formula *complete probabilities* (2.11): $P(B) = \frac{3}{5} \cdot 0.05 + \frac{1}{5} \cdot 0.03 + \frac{1}{5} \cdot 0.03 = 0.042$. That is a 4.2 percent chance that the stopped passenger is a smuggler. □

Let's now solve the inverse problem: if the control passenger is a smuggler, what is the probability that he is from the first city? I deliberately use such notations so that these examples can be easily generalized. The general answer is given by *Bayes's formula*, or the formula *probability hypothesis*.

If the events $A_1, A_2, ..., A_n$ form a single decomposition of Ω, then:

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{k=1}^{n} P(A_k)P(B|A_k)}, \quad i = 1, 2, ..., n. \tag{2.12}$$

The proof follows simply from $P(A_iB) = P(B)P(A_i|B) = P(A_i)P(B|A_i)$ from which, substituting $P(B)$ using the total probability formula, we obtain the required result.

Example 63. *What is the chance that the stopped passenger in the previous example is from the first city?*

Solution. In the previous, 62nd example, we had $P(B) = 0.042$ from $P(A_1) = 3/5$ and $P(B|A_1) = 0.05$ so the probability that a randomly stopped smuggler is from the first city is:

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(B)} = \frac{\frac{3}{5}\cdot 0.05}{0.042} = 0.714...$$

The chance is approximately 71.4 percent. □

The mentioned events can be something else (Diagnosis). There can be more of them and much more for computer work. For example, we monitor the import of drugs, bad alcohol, illegal weapons and the like as events $A_1, A_2, A_3, \dots$ with probabilities in the order $\alpha_1 = P(A_1)$, $\alpha_2 = P(A_2)$, $\alpha_3 = P(A_3)$ and so on. Let $s_{ij} = P(B_i|A_j)$ be the conditional probabilities that the cause A_j leads to the consequence B_i where $B_1, B_2, B_3, \dots$ are the occurrence of drug addiction, alcohol poisoning, murders and so on. Let the probabilities of these be $\beta_1 = P(B_1)$, $\beta_2 = P(B_2)$, $\beta_3 = P(B_3)$, ...

Drug trafficking does not necessarily lead only to an increase in drug addiction, but also to gun fights, the shadow economy, social corruption, and beyond, so many of these conditional probabilities s_{ij} can be positive numbers. Such criminal dependencies are observed by some services, and they have, maintain, and constantly improve matrices $S = (s_{ij})$. There are infinite possibilities of deviations, but based on the Borel-Cantelli lemma, we always have only one finite group of events that dominates.

The previous formulas (2.11) give $\beta_i = \alpha_1 p_{i1} + \alpha_2 p_{i2} + \alpha_3 p_{i3}$, in more detail:

$$P(B_i) = P(A_1)P(B_i|A_1) + P(A_2)P(B_i|A_2) + P(A_3)P(B_i|A_3), \quad i = 1, 2, 3$$

and all this can be written more briefly in matrix form:

$$\begin{pmatrix}\beta_1\\ \beta_2\\ \beta_3\end{pmatrix}\begin{pmatrix}s_{11} & s_{12} & s_{13}\\ s_{21} & s_{22} & s_{23}\\ s_{31} & s_{32} & s_{33}\end{pmatrix}\begin{pmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\end{pmatrix}, \tag{2.13}$$

or even shorter $b = Sa$. Here, the vectors are the arrays $a = (\alpha_1, \alpha_2, \alpha_3)$ and $b = (\beta_1, \beta_2, \beta_3)$, and $S = (s_{ij})$ is the matrix. Bayes' formula (2.12) gives $P(A_j|B_i) = s_{ij}/\beta_j$, the inverse conditional probability that it was the cause A_j that caused the effect B_j. It is clear how to generalize these equations to the order of matrices greater than three.

Let us now return to the probabilities of the product of events (2.10). Incidentally, we know that an impossible event is one with probability zero, so we say that an event is possible for any event that does not have probability zero. We say that disjoint events are mutually exclusive. If events are not independent, we say that they are dependent, and here is what it means exactly that events are "independent". I follow the book [10] to its unusual theorem 1.1.47, here the 67th theorem.

We say that an event A is *independent* of an event B when

$$P(A|B) = P(A). \tag{2.14}$$

If A is independent of B then B is also independent of A, because:

$$P(B|A) = \frac{P(BA)}{P(A)} = \frac{P(A|B)P(B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B).$$

Therefore, the independence of two events A and B can be defined by

$$P(AB) = P(A)P(B). \tag{2.15}$$

When we have multiple events $A_1, A_2, ..., A_n$, we say that they are independent if for each of their choices $A_{i_1}, ..., A_{i_m}$ we have

$$P(A_{i_1}...A_{i_m}) = P(A_{i_1})...P(A_{i_m}). \tag{2.16}$$

Unlike disjunction, which is defined using the events themselves, here we define independence using probabilities, measures on the events, so this "independence" should also be checked intuitively.

Example 64. *A regular tetrahedron has three sides colored in one of three different colors (red, green, and blue) and the fourth side is divided into three parts in those three colors. Check the independence of the color distribution.*

Solution. The events are: A the first color (red) is drawn, B the second color (green), C the third color (blue) is drawn. The probabilities are $P(A) = P(B) = P(C) = \frac{2}{4} = \frac{1}{2}$, $P(AB) = P(BC) = P(CA) = \frac{1}{4}$ and $P(ABC) = \frac{1}{4}$, but $P(ABC) \ne P(A)P(B)P(C)$, which means that these events are independent in pairs, but not in the whole. □

As a reminder, for Ω we say that a is a *sure event*, and for $\varnothing$ we say that is an *impossible event*.

Example 65. *If $A \subseteq \Omega$ is an arbitrary event, show that the following pairs of events are mutually independent:*

$$(A, \Omega), \quad (A, \varnothing).$$

Solution. From $A \cap \Omega = A$ and $P(\Omega) =$ it follows that $P(A \cap \Omega) = P(A) \cdot 1$, which is independence. Secondly, from $A \cap \varnothing = \varnothing$ and $P(\varnothing) = 0$ it follows that $P(A \cap \varnothing) = P(A) \cdot 0$, which also means independence. □

Example 66. *The pair of events (A, B) is independent. Show that they are independent:*

$$(A, B^c); \quad (A^c, B); \quad (A^c, B^c).$$

Solution. The events $A \cap B$ and $A \cap B^c$ are decompositions of the event A, so

$$P(A) = P(A \cap B) + P(A \cap B^c),$$

and given $P(A \cap B) = P(A)P(B)$. Hence these independences, respectively:

$$P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)[1 - P(B)] = P(A)P(B^c);$$

$$P(A^c \cap B) = P(B \cap A^c) = P(B)P(A^c) = P(A^c)P(B);$$

If (A, B) are independent, then (A, B^c) are independent. Again, if (B^c, A) are independent, then (B^c, A^c) are independent. □

Finally, we arrive at that famous theorem, which we will use for an unusual interpretation. Let's first look at what the theorem says.

Theorem 67. *Given two possible events A and B.*

1° *If the events A and B are independent, then they are not mutually exclusive.*

2° *If the events are mutually exclusive, then they are dependent.*

Proof. 1° For possible independent events A and B, assume the opposite, that they are mutually exclusive. Then $P(A)P(B) \neq 0$, because they are possible, but $P(A \cap B) = P(A)P(B)$, which is a contradiction to the assumption $P(A \cap B) = P(\varnothing) = 0$. Therefore, these events are not disjoint.

2° If the events A and B are mutually exclusive, they are disjoint, and then $P(A \cap B) = 0$, so $P(A|B)P(B) = 0$, and since $P(B) \neq 0$, it must be $P(A|B) = 0$, which means that the event A depends on B. □

The theorem states that for possible events (other than the empty set) the following equivalences hold:

$$\begin{cases} P(AB) = P(A)P(B) & \Leftrightarrow \quad P(AB) \neq 0, \\ P(AB) \neq P(A)P(B) & \Leftrightarrow \quad P(AB) = 0. \end{cases}$$

This means that "dependence" is equivalent to "exclusivity", and "independence" is equivalent to "nonexclusivity".

In a theory that assumes information as a weaving of space, time, and matter, and uncertainty as its essence, independent events are the basis of communication. What we get by exchanging information, which can be had in other ways, is not "news." When we know that something will happen, and it does happen, then it is not "novelty." Accordingly, *event independence* is at the heart of the weaving of reality in a way equivalent to non-exclusivity. Conversely, when we do not have independence, then information is absent, and without it, other phenomena of the world are extinguished. They are then excluded.

2.3.3 Conditional distribution

Consider matrices like (2.13). We have derived it as a scheme of conditional probabilities $s_{ij} = P(B_i|A_j)$, that the import of goods of type A_j leads to a social phenomenon B_i. As we consider it, it can represent individual symptoms that lead to certain diseases, or diseases that lead to therapies, or therapies that lead to certain financial, social, scientific, or natural phenomena. Mathematics is so practical, according to the *Löwenheim–Skolem* theorem, that everyone's perception, unlike subjects, is unique, which we take as an interpretation of the Riesz's theorem.

I remind you that Riesz's theorem in the analysis (mentioned link) says that the bounded linear functional $f : X \to \Phi$, which is a mapping of the vector space ℓ_1 into scalars, has the representation:

$$f(x) = \sum_{\nu=1}^{\infty} \eta_\nu \xi_\nu, \quad x = (\xi_\nu), \ y = (\eta_\nu) \in m, \ \|f\| = \|y\|_m,$$

and the functional f on ℓ_1 corresponds to one and only one point y in m.

This theorem is also valid for other spaces, for example x-spaces from p-norm when y from it is dual to q-norm ($1/p + 1/q = 1$, $p \ge 1$), where here ℓ_1 is of norm $p = 1$, and its dual is m of norm $q = \infty$. I chose this one:

$$\|x\|_1 = |\xi_1| + |\xi_2| + |\xi_3| + \dots, \quad \|y\|_\infty = \max_{1\le\nu<\infty}\{|\eta_\nu|\},$$

when the sum $\|x\|_1$ converges, because that sum is exactly the sum of probabilities. Now, these x-spaces are vectors of norm $\|a\| = |\alpha_1| + |\alpha_2| + \dots + |\alpha_N|$, or $\|b\| = |\beta_1| + |\beta_2| + \dots + |\beta_M|$, and also $\|s_{i.}\| = |s_{i1}| + |s_{i2}| + \dots + |s_{iN}|$ respectively for $i = 1, 2, \dots, N$, where M is the number of rows of the matrix S, and N is the number of its columns. When each of these sums is 1, then we are working with *probability distributions*, and the matrix is square ($M = N$), then we are also working with *stochastic matrices*.

In general, especially in the theory of information perception [7], I believe that the Riesz theorem establishes that for a way of perceiving the world ($x \in X$) there is one and only one state, a subject (dual vector $y \in X$) that perceives, communicates, or experiences the world around it in a unique way. There is no other subject in the entire universe that has that same representation, and it is not always about some large, visible differences. The story is completely opposite about theories, universal truths that last in such a way that they never change for the mentioned subjects. These subjects are only their own, and those theories are everyone's – I say sometimes, referring to what has just been explained.

A stohastic matrix is a square matrix whose columns are probability vectors. A probability vector is a numerical vector whose entries are real numbers between 0 and 1 whose sum is 1. Stochastic matrices are generators, links of a *Markov chain* that, by composing such mappings, therefore cascades probability distributions and is then called a Markov matrix.

Example 68. *The stochastic matrix translates a distribution into a distribution.*

Proof. From equation (2.13) we take:

$$\beta_i = s_{i1}\alpha_1 + s_{i2}\alpha_2 + s_{i3}\alpha_3, \quad \beta_i \ge 0,$$

because all factors are assumed to be non-negative. Further, the sum is:

$$\sum_{i=1}^{3}\beta_i = \sum_{i=1}^{3}\sum_{j=1}^{3} s_{ij}\alpha_j = \sum_{j=1}^{3}\left(\sum_{i=1}^{3} s_{ij}\right)\alpha_j = \sum_{j=1}^{3} 1\cdot\alpha_j = 1.$$

The sum in parentheses is 1, because it is the sum of the j-th column of the stochastic matrix, and the last sum is 1, because a is a distribution. Therefore, b is also a distribution. □

The example given is a stochastic matrix 3×3 so that you can more easily "brute force"the proof by printing out all the sums, but it certainly stands to reason that this $n = 3$ can be changed with the same effect. To be precise, we say that this is a "left stochastic"matrix, a square matrix of non-negative real numbers whose

column sum is 1. It stands to the left of the column vector it multiplies. The second is a "right stochastic"matrix, also a square matrix of non-negative real numbers, but whose row sum is 1. It stands to the right of the species vector it multiplies. The third is a đouble stochastic"matrix, again a square matrix of non-negative real numbers, but with each row and each column sum being 1.

Note that the left and right stochastic matrices behave as if they were adjoints (transposed) of each other, and then the doubly stochastic is self-adjoint (symmetric, because it has real coefficients). Therefore, all eigenvalues of λ of the doubly stohastic matrix, hence the solutions of its characteristic equation $Sv = \lambda v$, are real numbers.

Theorem 69. *The largest eigenvalue of a stohastic matrix is 1.*

Proof. If A is a stochastic matrix, then from $Av = \lambda v$ it follows that $(A - I)v = 0$ and $\det(A-I) = 0$ due to the unity sum of the columns. Therefore, $\lambda = 1$ is an eigenvalue of every stochastic matrix.

Suppose further that there exists $|\lambda| > 1$ and a vector $x \neq 0$ such that $Ax = \lambda x$. Let ξ_i be the largest element of x. Since any scalar multiple of x will also satisfy this equation, we can assume, without loss of generality, that $\xi_i > 0$. Since the columns of the matrix A are nonnegative integers with sum 1, every entry in $|\lambda|x$ is a convex combination of the elements of x . Thus, no entry in $|\lambda|x$ cannot be greater than ξ_i. But since $|\lambda| > 1$, it follows that $|\lambda|\xi_i > \xi_i$, which is a contradiction. So, the largest eigenvalue of A is 1. □

The eigenvector v at eigenvalue 1 is called the stable equilibrium distribution of the matrix A. It is also called the Perron-Frobenius eigenvector.

Example 70. *The product of stochastic matrices is the stochastic matrix*[4].

Proof. Let's look at the multiplication of $P = QS$ matrices $P(p_{ij})$, $Q(q_{ij})$ and $S(s_{ij})$ and the sum of the elements of the i-th column of the result:

$$\sum_{j=1}^{3} p_{ij} = \sum_{j=1}^{3}\sum_{k=1}^{3} q_{ik}s_{kj} = \sum_{k=1}^{3} q_{ik}\left(\sum_{j=1}^{3} s_{kj}\right) = \sum_{k=1}^{3} q_{ik}\cdot 1 = 1,$$

since by assumption Q and S are stochastic matrices, so P is stochastic. □

Task 71. *Find the eigenvalues and vectors of the stochastic matrix:*

$$A = \begin{pmatrix} 0.7 & 0.1 & 0.2 \\ 0.2 & 0.6 & 0.3 \\ 0.1 & 0.3 & 0.5 \end{pmatrix}.$$

Solution. We solve the characteristic equation $Av = \lambda v$, i.e. $(A - \lambda I)v = 0$:

$$\det\begin{pmatrix} 0.7-\lambda & 0.1 & 0.2 \\ 0.2 & 0.6-\lambda & 0.3 \\ 0.1 & 0.3 & 0.5-\lambda \end{pmatrix} = 0.$$

[4] We assume left stochastic.

Let's add the second and third rows to the first, and then subtract the first from the second and third columns. We get the determinant, which we expand along the first line:

$$\begin{vmatrix} 1-\lambda & 0 & 0 \\ 0.2 & 0.4-\lambda & 0.1 \\ 0.1 & 0.2 & 0.4-\lambda \end{vmatrix} = 0,$$

$$(1-\lambda)\begin{vmatrix} 0.4-\lambda & 0.1 \\ 0.2 & 0.4-\lambda \end{vmatrix} = 0,$$

$$(1-\lambda)[(0.4-\lambda)^2 - 0.02] = 0,$$

$$\lambda_1 = 1, \quad \lambda_{2;3} = 0.4 \pm \sqrt{0.02}.$$

The eigenvalues are $\lambda_1 = 1$, $\lambda_2 \approx 0.54142$ and $\lambda_3 \approx 0.25858$.

We calculate the first, second and third eigenvectors ($Av = \lambda v$):

$$\begin{cases} (0.7-\lambda)x + 0.1y + 0.2z = 0 \\ 0.2x + (0.6-\lambda)y + 0.3z = 0 \\ 0.1x + 0.3y + (0.5-\lambda)z = 0 \end{cases}$$

$$(-9+20\lambda)y = (-17+30\lambda)x, \quad (1+10\lambda)z = (-20+30\lambda)x.$$

We include the eigenvalues λ_1, λ_2 and λ_3. For the first one we find:

$$v_1 = \frac{1}{34}\begin{pmatrix} 11 \\ 13 \\ 10 \end{pmatrix},$$

The sum of the elements of the first eigenvector v_1 is 1 and this is the probability distribution. The other eigenvectors do not have to be probabilities. Moreover, they can be negative, even complex numbers. □

Here are a few more interesting things, not to say anomalies, that can "surprise" us when solving the characteristic equation $Av = \lambda v$ of the stochastic matrix A. The following three stochastic matrices, in order:

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix},$$

show that they can have multiple unit eigenvalues, that they can have negative eigenvalues, and that they exist with an eigenvalue as well as a determinant of zero. The following demonstrates complex eigenvalues:

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$

because the zeros of the polynomial $\lambda^3 - 1 = 0$ are $\lambda_1 = 1$ and $\lambda_{23} = (-1 \pm \sqrt{3}i)/2$. It is also an example of the "deviation" when the stohastic matrix is orthogonal.

For a general $n \times n$ matrix $A = (a_{ij})$ the following *Gershgorin's theorem* (1931) defines n disks (circles) of the complex plane whose union contains the eigenvalues of the matrix. We use it for various evaluations. The Gershgorin disks form the set $\{D_1, D_2, ..., D_n\}$ each with a corresponding radius:

$$R_i = \sum_{j \neq i} |a_{ij}|, \quad D_i = \{z \in \mathbb{C} \ : \ |z - a_{ij}| \leq R_i\}.$$

Theorem 72. *Let $A = (a_{ij})$ be an $n \times n$ matrix with real or complex coefficients. It is eigenvalues are in Gershgorin disks.*

Proof. Let $x = (\xi_1, \xi_2, ..., \xi_n) \neq 0$ so that $Ax = \lambda x$. Let ξ_i be the largest modulus coefficient of the sequence x. We have:

$$Ax = \lambda x \Rightarrow \sum_{j=1}^{n} a_{ij}\xi_j = \lambda \xi_i \Rightarrow \sum_{j \neq i} a_{ij}\xi_j = (\lambda - a_{ij})\xi_i$$

hence:

$$|\lambda - a_{ii}| = \left| \sum_{j \neq i} \frac{a_{ij}\xi_j}{\xi_i} \right| \leq \sum_{j \neq i} |a_{ij}| = R_i.$$

□

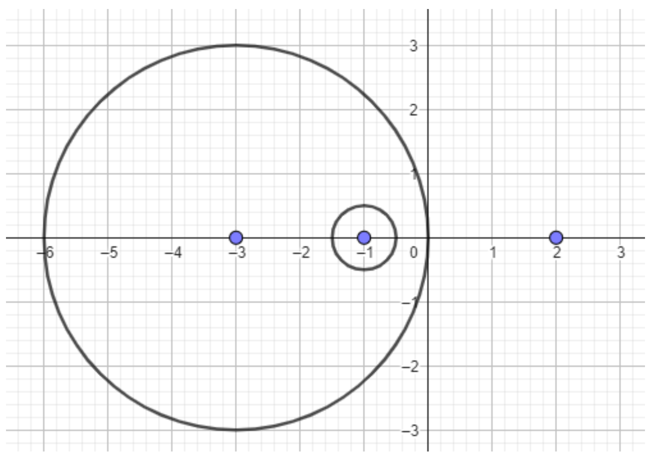
For example, given the matrix:

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0.5 \\ -2 & 1 & -3 \end{pmatrix}.$$

It has Gershgorin disks:

$$D_1 = \{z \in \mathbb{C} \ : \ |z - 2| \leq 0\}, \quad D_2 = \{z \in \mathbb{C} \ : \ |z + 1| \leq 0.5\}, \quad D_3 = \{z \in \mathbb{C} \ : \ |z + 3| \leq 3\}$$

shown in the following complex plane figure.



Inside these disks are the eigenvalues of the given matix, so because of the zero diameter D_1 we know $\lambda_1 = 2$. The second (λ_2) is inside a circle of small range (radius

$R_2 = 0.5$) centered at -1. The third (λ_3) is only roughly estimated around the point -3 of the larger radius ($R_3 = 3$).

The matrix and the transpose matrix have the same eigenvalues, so the Gershgorin disks also apply when we take the column elements of the given matrix. Then:

$$D_1' = \{z \,:\, |z| \le 2\}, \quad D_2' = \{z \,:\, |z+1| \le 1\}, \quad D_3' = \{z \,:\, |z+3| \le 0.5\}.$$

This can further improve the estimation of the eigenvalues. For example, the third eigenvalue (λ_3) is in the complex number range around the point -3 to the radius $R_3 = 0.5$. Otherwise, for the given matrix, $\lambda_1 = 2$, $\lambda_2 \approx -1.29289$, and $\lambda_3 \approx -2.70711$.

When we talk about the application of Gershgorin's theorem through Markov chains and stochastic matrices, we note that diagonal coefficients larger than 0.5 guarantee the transfers of the k-th components of the distribution into even larger k-th components. Conditional probabilities s_{ij} are a measure of the transfer of $j \to i$, when $j = i$ means the correct transfer, and $j \ne i$ the transfer error, or disinformation. In addition, since the coefficients of these matrices are from the interval of numbers (0, 1), diagonals larger than 0.5 guarantee that there are no zero eigenvalues, which means that the determinant of the matrix is not zero, i.e., that the input can be read based on the output message.

2.3.4 Markov chain

By definition, the determinant of a matrix (linear operator) is the product of its eigenvalues. Therefore, the determinant is zero only when one of the eigenvalues is zero. On the other hand, the determinant of order n is a measure of the generalized volume of the n-dimension parallelepiped spanned by its column vectors. This means that its column vectors are linearly dependent if one of the eignvalues is zero.

In other words, if a matrix (operator) reduces a vector space to less than n dimensions, when the copy loses at least one of the dimensions of the original and therefore the matrix cannot have an inverse, its determinant is zero. At least one of its eigenvalues is then zero. Since dependent events are mutually exclusive (Theorem 67), they do not communicate because the "news" that we know in advance, or is certain in any way, is, to put it mildly, not news. Since we consider action and information to be equivalent, then there is no transfer of interaction, there is no corresponding observable, and therefore some of the eigenvalues is zero.

This is a brief computer science interpretation of the well-known theorems of linear algebra. We will now see how much it helps in understanding the cascading transfer of information. We observe a discrete (stepwise) message flow through a Markov chain, with multiple use, that is, the composition of an arbitrary and still constant stochastic matrix, which we call the chain generator:

$$S = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1n} \\ s_{21} & s_{22} & \dots & s_{1n} \\ \dots & & & \\ s_{n1} & s_{n2} & \dots & s_{nn} \end{pmatrix}.$$

The initial, zero message is the probability distribution $x_0 = (\xi_{01}, \xi_{02}, ..., \xi_{0n})$. The second message $x_1 = Sx_0$ is the first transmitted $x_1 = (\xi_{11}, \xi_{12}, ..., \xi_{1n})$. It is also some probability distribution (example 68). The transmitted one gives the next $x_2 = Sx_1 = S^2x_0$, so if we continue through the chain cascaded, in the k-th step $x_k = Sx_{k-1} = S^kx_0$ there will be a probability distribution $x_k = (\xi_{k1}, \xi_{k2}, ..., \xi_{kn})$. We ask questions about finding x_0 based on x_k, in order for $k = 1, 2, 3, ...$.

Typical contemporary information transmitters are the telegraph, telephone, radio, television, designed to serve our communication, but the theory of *Markov chain* far exceeds these ranges. Therefore, here we consider these "typical" stochastic matrices to be those that do not have zero eigenvalues and for which they are not all the same. In other words, a typical Markov chain is one that will give readable messages, but not indefinitely.

Namely, if in the characteristic equation $Sx = \lambda x$ each $\lambda \ne 0$, then $\det S \ne 0$, and this has the consequence of the readability of the copy based on the original. However, again from $\det S = \lambda_1\lambda_2...\lambda_n$, because this product is a lambda number modulo less than one (theorem 69) it will be:

$$\lim_{k\to\infty} \det S^k = \lim_{k\to\infty} (\det S)^k = 0,$$

because the determinant of the product is equal to the product of the determinants, and the limit value of the degree of a number modulo less than one is zero. Thus, a typical Markov chain becomes atypical after many links, in the sense that we cannot read the initial message from the final message. We know the algebra of this (example 50), but perhaps we don't notice the breadth of its application.

Example 73. *Let's say a company has $n = 3$ locations in our city where it lends goods (bicycles, movies, money, etc.). What is returned at the i location can be rented at the j location, so the user statistics are formed:*

$$A = \begin{pmatrix} 0.3 & 0.4 & 0.5 \\ 0.3 & 0.4 & 0.3 \\ 0.4 & 0.2 & 0.2 \end{pmatrix}.$$

This is a stochastic matrix with about 50% of the funds borrowed at the third location and returned at the first location, or $a_{32} = 20$ percent borrowed at the $j = 2$ location and returned at $i = 3$.

The eigenvalues ($Av = \lambda v$) of this matrix are $\lambda_1 = 1$, $\lambda_2 = -0.2$ and $\lambda_3 = 0.1$, with the corresponding eigenvectors:

$$v_1 = \begin{pmatrix} 7 \\ 6 \\ 5 \end{pmatrix}, \quad v_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix},$$

where we can divide the first by $7 + 6 + 5 = 18$, if we want it to be a probability distribution. In the process of diagonalizing the matrix, we mentioned the auxiliary

matrix $P = P[v_1, v_2, v_3]$ whose columns are such eigenvectors, and next to it is the inverse matrix ($PP^{-1} = P^{-1}P = I$):

$$P = \begin{pmatrix} 7 & -1 & 1 \\ 6 & 0 & -3 \\ 5 & 1 & 2 \end{pmatrix}, \quad P^{-1} = \frac{1}{18}\begin{pmatrix} 1 & 1 & 1 \\ -9 & 3 & 9 \\ 2 & -4 & 2 \end{pmatrix}.$$

If the eigenvalues are different from each other, then the eigenvalues are linearly independent and there is an inverse matrix. Then its columns can be represented in the standard basis ($D = P^{-1}AP$) when the given matrix A is diagonalized:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -0.2 & 0 \\ 0 & 0 & 0.1 \end{pmatrix} = \frac{1}{18}\begin{pmatrix} 1 & 1 & 1 \\ -9 & 3 & 9 \\ 2 & -4 & 2 \end{pmatrix}\begin{pmatrix} 0.3 & 0.4 & 0.5 \\ 0.3 & 0.4 & 0.3 \\ 0.4 & 0.2 & 0.2 \end{pmatrix}\begin{pmatrix} 7 & -1 & 1 \\ 6 & 0 & -3 \\ 5 & 1 & 2 \end{pmatrix}.$$

Represented using the diagonal, the given matrix is easily exponented. Namely, from $A = PDP^{-1}$ it follows:

$$A^k = (PDP^{-1})^k = (PDP^{-1})(PDP^{-1})...(PDP^{-1}) =$$

$$= PD(P^{-1}P)D(P^{-1}...P)DP^{-1} = PD^kP^{-1}.$$

However, it is easy to rank a diagonal matrix, because D^k again diagonal elements on the diagonal of the rank of the corresponding elements of the matrix D. Now we get the degree of the given matrix A^k by multiplying only three matrices:

$$\begin{pmatrix} 0.3 & 0.4 & 0.5 \\ 0.3 & 0.4 & 0.3 \\ 0.4 & 0.2 & 0.2 \end{pmatrix}^k = \begin{pmatrix} 7 & -1 & 1 \\ 6 & 0 & -3 \\ 5 & 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & (-0.2)^k & 0 \\ 0 & 0 & (0.1)^k \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 \\ -9 & 3 & 9 \\ 2 & -4 & 2 \end{pmatrix}\frac{1}{18}.$$

After a long number of steps, $k \to \infty$, the powers of the numbers modulo less than one become zero, so we have:

$$\begin{pmatrix} 0.3 & 0.4 & 0.5 \\ 0.3 & 0.4 & 0.3 \\ 0.4 & 0.2 & 0.2 \end{pmatrix}^k \to \begin{pmatrix} 7 & -1 & 1 \\ 6 & 0 & -3 \\ 5 & 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 \\ -9 & 3 & 9 \\ 2 & -4 & 2 \end{pmatrix}\frac{1}{18} =$$

$$= \begin{pmatrix} 7 & -1 & 1 \\ 6 & 0 & -3 \\ 5 & 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\frac{1}{18} = \begin{pmatrix} 7 & 7 & 7 \\ 6 & 6 & 6 \\ 5 & 5 & 5 \end{pmatrix}\frac{1}{18} = A^\infty.$$

This boundary matrix (A^∞) is also stochastic, but with equal columns v^∞ = (7/18, 6/18, 5/18) which do not span the 3-dimensional space, like the initial (A), so it does not have an inverse matrix. It transforms every distribution $x = (\xi_1, \xi_2, \xi_3)$ into v^∞, which is easy to check by multiplication ($A^\infty x = v^\infty$). We call it a *black box*.

A long accumulation of abandoned goods, with steps of the matrix A, becomes a stable statistical distribution v^∞ when about 39 percent ($100 \cdot 7/18$) of them end up in the first position, 33 percent ($100 \cdot 6/18$) in the second, and 28 percent ($100 \cdot 5/18$)

in the third, regardless of where the funds were taken from. This is the process of small-to-large sample growth, which is affected by the *law of large numbers* of probability theory.

There are other applications, or understandings, of the same result, some of which we will return to later, but there are also other ways of computing this result, some of which we can look at right away. First of all, because the eigenvectors of a matrix A of order $n = 3$ span a vector space of the same dimension, it is possible to represent any vector in that space by them. This is the procedure for solving the following problem.

Task 74. *Find the "black box"of the stochastic matrix of the previous example, representing the general vector by a linear combination of the eigenvalues.*

Solution. The matrix is

$$A = \begin{pmatrix} 0.3 & 0.4 & 0.5 \\ 0.3 & 0.4 & 0.3 \\ 0.4 & 0.2 & 0.2 \end{pmatrix},$$

its eigenvalues are $\lambda_1 = 1$, $\lambda_2 = -0.2$ and $\lambda_3 = 0.1$ of the corresponding ($Av = \lambda v$) eigenvectors:

$$v_1 = \frac{1}{18}\begin{pmatrix} 7 \\ 6 \\ 5 \end{pmatrix}, \quad v_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix}.$$

The eigenvalues are distinct and the eigenvectores are linarly independent. They can be used to represent an arbitrary vector $x = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3$ of the same 3-dimensional space. Applying a linear operator to a vector:

$$Ax = A(\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3) = \alpha_1 Av_1 + \alpha_2 Av_2 + \alpha_3 Av_3 = \alpha_1\lambda_1 v_1 + \alpha_2\lambda_2 v_2 + \alpha_3\lambda_3 v_3,$$

$$A(Ax) = A^2 x = \alpha_1\lambda_1^2 v_1 + \alpha_2\lambda_2^2 v_2 + \alpha_3\lambda_3^2 v_3,$$

$$A(A^2 x) = A^2 x = \alpha_1\lambda_1^3 v_1 + \alpha_2\lambda_2^3 v_2 + \alpha_3\lambda_3^3 v_3$$

and after $k = 1, 2, 3, \dots$ steps

$$A^k x = \alpha_1\lambda_1^k v_1 + \alpha_2\lambda_2^k v_2 + \alpha_3\lambda_3^k v_3,$$

$$A^k x = \alpha_1 v_1 v_1 + \alpha_2(-0,2)^k v_2 + \alpha_3(0,1)^k v_3,$$

$$\lim_{k\to\infty} A^k x = \alpha_1 v_1 = \frac{\alpha_1}{18}\begin{pmatrix} 7 \\ 6 \\ 5 \end{pmatrix}.$$

Only the coefficient with the leading eigenvalue ($\lambda_1 = 1$) is important. Hence:

$$\lim_{k\to\infty} A^n = \frac{1}{18}\begin{pmatrix} 7 & 7 & 7 \\ 6 & 6 & 6 \\ 5 & 5 & 5 \end{pmatrix} = A^\infty.$$

This (A^∞) is the required “black box”. □

When we take a closer look at the solution to this problem, we notice that the output vector in the limiting case ($k \to \infty$) completely takes the direction of the leading eigenvector (v_1) of the mapping, the matrix A. In particular, the other degrees ($\lambda^k \to 0$) give less and less importance to this result, which means that an arbitrary message through Markov links becomes closer and closer to the leading eigenvector. The processes of such approximation are adaptations. The consequences are the principles of minimalism: spontaneous tendencies to less action, less information, more probable outcome.

Real processes under steady-state conditions may be close to Markovian. However, because of the duration of "reality" (subject to conservation laws), we must look at the processes backwards, from the present and by steps back into the past. The "black box" is then a period, a sequence, or a composition of mappings from some distant past event to the present, which under the above conditions should depict an amorphous, impersonal beginning.

Assuming that we have such an approximate regularity in the development of the universe, or that sequences with slightly changed Markov chain generators could resemble the above, then the Big Bang (the origin of the universe 13.8 billion years ago) could resemble it. This would mean that what we see as a hot, shapeless soup is not exactly what it was then, and that we may never establish a "better" truth than one that would be above this, mathematical one. All observations, measurements, and the laws of experimental sciences based on them are subordinate to mathematical ones in the sense that they cannot be used to prove or disprove even the Pythagorean theorem.

2.3.5 Canal noise

Markov chains of a typical generator (with various and non-zero eigenvalues) have transmission noise, *channel noise*, simply because a longer composition constantly produces new transmitters that make different copies. We are now interested in the structure of these changes.

For example, the mean value of the information taken in the Shannon way, the initial matrix itself and the black box of the previous example is:

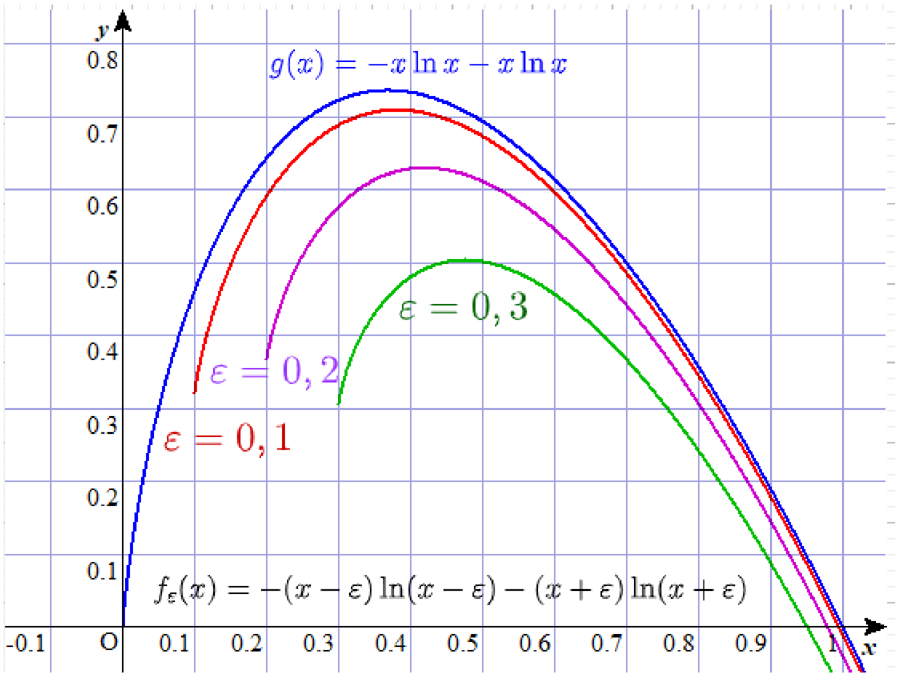
$$\text{Inf}(A) = -0.3 \cdot \ln 0.3 - 0.4 \cdot \ln 0.4 - ... - 0.2 \cdot \ln 0.2 \approx 3.17347$$

$$\text{Inf}(A^\infty) = -3 \cdot \frac{7}{18} \cdot \ln \frac{7}{18} - 3 \cdot \frac{6}{18} \cdot \ln \frac{5}{18} - 3 \cdot \frac{5}{18} \ln \frac{7}{18} \approx 3.26793.$$

This increase in information (from about 3.17 to about 3.27) with the length of the composition indicates that the interference increases the information of the channel. Distractions are disinformation and they actually spice up, burden the incoming message, making it less and less recognizable.

The appearance of noise results from changes in the conditional probabilities of the composition of the matrices (A^k, $k = 1, 2, 3, ...$) that make them uniform. The uniform distribution has more information simply because the more probable outcome is less informative, that is, because the less predictable news is more informative. That the calculation of the mean, in the manner of Shannon information,

otherwise confirms this intuitive expectation in the general case – we establish below. I do this in steps to emphasize some new observations.



Example 75. *Unequal probability distributions, and even their parts, have lower Shannon information than equal ones.*

Explanation. The above figure shows the graphs $f_\varepsilon(x) < g(x)$. The top one is blue, $g(x) = -2x \cdot \ln(x)$, and below are the graphs of the functions:

$$f_\varepsilon(x) = -(x-\varepsilon)\cdot\ln(x-\varepsilon) - (x+\varepsilon)\cdot\ln(x+\varepsilon)$$

red, purple and green, on intervals $x \in (\varepsilon, 1-\varepsilon)$. It can be seen that in the case of Shannon information:

$$\dots + f_\varepsilon(x) + \dots < \dots + g(x) + \dots,$$

where their unwritten sums are equal. Left written sums, the function $f_\varepsilon(x)$ with increasing $\varepsilon \in \{0,1; 0.2; 0.3\}$ the lower the graph, the more inhomogeneous the decompositions of the right $g(x)$ are. This shows: the less uniform the probability distribution, the lower the Shannon information. □

The lesson of the example is “nature does not like equality”. I paraphrase the principle of minimalism: a system spontaneously tends to a state of less information or higher probabilities. We see it in the principle of least action that is the foundation of theoretical physics. However, the principle of minimalism also operates, say, in *network theory* (of equal links) whose nodes at the ends of equal links tend to form so that few of them have many links compared to many others. This

creates fewer degrees of separation between nodes and makes their communication cheaper.

For example, a free market, when it means the equal circulation of money, goods and services, with the owners as nodes, will spontaneously develop into a small number of owners of a lot and the rest of the population poorer. We believe that with money comes the power of the owner, but for this theory this is secondary. Certainly, under conditions of equality of rights and power, a few who have more will spontaneously emerge above the multitude of others.

The separation of galaxies into a kind of island of the growing universe was similarly initiated, or the emergence of the laws of physics for, say, the formation of atoms as a kind of concentration within substances. I am talking about the processes of the cosmic amorphous initial soup of 13 billion years ago. In the early universe, the electromagnetic and weak forces were one, and over time they separated and created the electroweak force. But as the universe began to cool, the weak nuclear force became isolated, and the strong force appeared alongside it.

However, the formation of only one node with all links as opposed to all other nodes with one each, in systems of equal graphs, will not happen spontaneously for the same reason – because then there would be too many equal others.

For this problem, the function $h(x) = x - 1 - \ln x$ whose increasing derivative $h'(x) = 1 - 1/x$ with zero at $x = 1$ shows that $h(1) = 0$ is its minimum. This means that

$$\ln x < x - 1$$

everywhere where the expression is defined, except at the point $x = 1$ when the equality holds. This inequality is graphically proven and then used in the appendix Extremes, as follows. Substituting $x = q_k/p_k$ into the inequality and multiplying it by p_k, respectively for $k = 1, 2, ..., n$ two probability distributions $p_1, p_2, ..., p_n$ and $q_1, q_2, ..., q_n$. Adding, we get:

$$\sum_{k=1}^{n}(p_k \ln q_k - p_k \ln p_k) < \sum_{k=1}^{n}(p_k - q_k) = 1 - 1 = 0,$$

$$-\sum_{k=1}^{n} p_k \ln p_k < -\sum_{k=1}^{n} p_k \ln q_k. \tag{2.17}$$

First of all, we see that combining the information ($-\ln q_k$) and the probability (p_k) in the "mean value"of the right-hand side of the inequality gives a larger value than the left-hand side where the same probabilities appear. Note that this inequality holds for any choice of the second distribution (q_k) and becomes an equality only when each $p_k = q_k$.

Example 76. *A uniform discrete probability distribution has maximum Shannon information.*

Proof. For a uniform distribution, $q_k = 1/n$, for each $k = 1, 2, ..., n$, its Shannon information is:

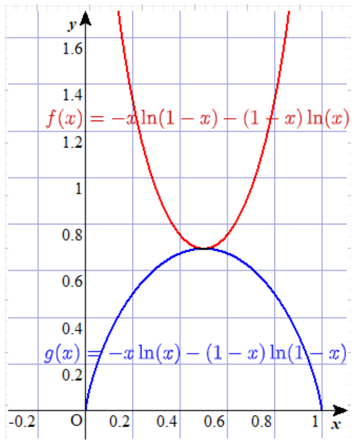
$$-\sum_{k=1}^{n} \frac{1}{n} \ln \frac{1}{n} = \ln n.$$

Substituting it into the logarithm of the right-hand side of inequality (2.17), we find:

$$-\sum_{k=1}^{n} p_k \ln p_k < \ln n,$$

where the equality holds if and only if $p_k = 1/n$ for each $k = 1, 2, ..., n$. □

In the figure on the right we see the graphs of two functions, the lower Shannon information g of the binomial probability distribution $p = x$ and $q = 1 - x$, and the upper f when the information is swapped:



$$f(x) = -x\ln(1-x) - (1-x)\ln(x),$$

$$g(x) = -x\ln(x) - (1-x)\ln(1-x).$$

It is visually clear why $g(x) < f(x)$ is always and why the inequality becomes an equality only if $x = 0.5$.

Here is an analytical proof. The inequality holds:

$$(-p+q)\ln\frac{p}{q} < 0,$$

whenever $p > q$ (the first factor is negative $-p + q < 0$ and the second is positive), or $p < q$ (the first factor is positive and the second is negative). But if $p = q$ then the equality holds. Further we have:

$$-p\ln\frac{p}{q} + q\ln\frac{p}{q} < 0,$$

$$-p(\ln p - \ln q) + q(\ln p - \ln q) < 0,$$

$$-p\ln p - q\ln q < -p\ln q - q\ln p,$$

so putting $p = x$ and $q = 1 - x$ we get the functions from the graph and their inequality $g(x) < f(x)$. Continuing as in the previous, 75th example, let us consider the Shannon information $S = -s_1 \ln s_1 - s_2 \ln s_2 - ... - s_n \ln s_n$, where s_k are the probability distributions. Whenever two sums exchange places in the manner of the above inequality, this sum increases.

The inequality of discrete distributions (2.17) is easily formulated for continuous ones:

$$-\int_a^b f(x)\ln f(x)\,dx \le -\int_a^b f(x)\ln g(x)\,dx. \tag{2.18}$$

Start from the same previous logarithmic inequality, swap the corresponding densities, and replace integration with addition. Here the equality is valid only if $g(x) = f(x)$ at all but a discrete number of points. On the left of the inequality is the Shannon information density of the probability distribution $f(x)$, and on the right we have the combination of those with the densities $g(x)$.

Probability density is in general a function $\rho(x) \geq 0$ that is integrable on the interval $x \in (a,b) \subseteq (0,1)$ and such that:

$$\int_a^b \rho(x)\,dx = 1.$$

In inequality (2.18) both functions $f(x)$ and $g(x)$ are (different) densities on the same interval. Corresponding to the previous one is the following example.

Example 77. *The uniform continuous probability density distribution $g(x)$ has a Shannon information $\ln(b-a)$ which is maximal for all densities of the interval $x \in (a,b)$, where $0 \leq a \leq b \leq 1$.*

Proof. The uniform continuous density is $g(x) = 1/(b-a)$, $x \in (a,b) \subseteq (0,1)$. Its Shannon information is:

$$-\int_a^b g(x)\ln g(x)\,dx = \int_a^b \frac{\ln(b-a)}{b-a}\,dx = \ln(b-a).$$

Putting it into inequality (2.18) gives:

$$-\int_a^b f(x)\ln f(x)\,dx \leq \ln(b-a),$$

which is what is claimed in this example. □

So, for any probability density distribution $\rho(x)$ on a given interval, it will be:

$$-\int_a^b \rho(x)\ln\rho(x)\,dx \leq \ln(b-a). \tag{2.19}$$

This inequality is applicable if we work with continuous operators, Markov chain generators, instead of discrete ones and their matrices.

The second lesson of these findings is that “misinformation hides information”. When we think about the transmission of information, most of us think of a device or a job where work is put into pushing messages from one place to another. In a typical[5] Markov chain, we spend media and similar effort and the information of the outgoing message grows with the noise. But overall, with its misinformation, it drowns out useful information for us, which becomes less and less readable to distant recipients. Let’s not forget that errors also occur when reading with the latest memory technologies (ssd, hdd, ram) of modern computers.

The opposite happens when we “read” messages from the past from the present. It could be some archaeological research, forensic analysis, astronomical observation of deep space, or something similar. We stand at the end, as if at the beginning of a Markov chain – which, from its ever-more distant links, tells us what was then. And the further “then” goes, the more and more burdened with the noise of

[5]typical – message is readable, but noise grows

the “channels”. That “message” that is closer to us in time becomes clearer, more definite, with less and less uncertainty, unburdened by “misinformation” and in that sense less informative.

According to this, the kind of channel noise in the transmission of information will also be a geometric perspective. It makes the objects of observation smaller and smaller the further they are, and thus participating in this theory and on the side of the conclusions about the disturbances of typical Markov chains. We know that space-time is the whole of relativistic physics, so this mixing of "purešpace into time processes should not be strange to us. This encourages us to consider the boundary of the visible universe, the *event horizon*, from that point of view, for example, as a *black box* of a Markov chain.

The other cosmic one, now I will call it the "event horizon", better known as the *Big Bang* of about 13.8 billion years ago, I have already mentioned as a "black box". Information has traveled such a long Markov chain up to the present that it has *completely clogged* with channel noise and now tells us that the cosmos was then a boiling amorphous soup. Since it is a mathematical truth, it is a kind of "reality"that cannot be challenged by some contradiction, a different mathematical "truth", and especially not by some methods of experimental science. This world of đeception of reality"is so perfect.

The more familiar event horizon is the boundary of a “black hole,” a sphere around a celestial body of such immense mass that not even light can escape it. This phenomenon also fits into the Marvel chain as a “black box” because time slows down the closer a point is to the center of mass. From the perspective of an outside observer, such as us, it takes an infinite amount of time for a rock to fall into a black hole. It also takes that long for information to travel from it to us. With these few discoveries about real-world illusions, these topics, new theories of information, that await us are nowhere near exhausted.

2.3.6 Defiance

All known trajectories to date have been successfully reduced by theoretical physics to the *principle of least action*. For example, in the book Minimalism of Information [12] (2.5 Einstein’s General Equations), you will find the derivation of the equations of general relativity from this principle, or in the appendix on light here (example 2), how the path of reflected light is the shortest. The discovery of these minimalists of quantum mechanics tells us that they do not have to be unambiguous, but not even Brownian motion nor any random motion (Physical Information [11], 2.5.2 Free walk of a point) can simulate, say, the motion of an ant.

Living beings have a complement of freedoms of choice that I call for now *vitality*, which dead physical nature cannot have simply because it cannot lie, does not notice lies, and does not react to them. Due to the absence of lies, at least in proving truths, a consistent theory (without falsehoods) is incomplete (not complete). This is the content of Gödel’s discoveries, with which I also explain the differen-

ce between mathematics, which uses incorrect assumptions, and physical reality without this possibility (1.3.3 Incompleteness). Due to the lack of these vitality options, inanimate beings are condemned to the principle of least action, and all others will tend towards it with the principle of minimalism.

In other words, living beings can *defy* the slightest action by avoiding it through manipulation. However, a lie costs; it is an "investment" that resists the spontaneity of nature, its forces against which we must fight with our vitality and which will tire us. By moving with this added will, acting first of all on our physical body and then on the environment, we really manage to do things that nature itself is not capable of. We are that disruptive factor that disrupts natural consistency in favor of something complete.

As we have already seen, the possibility of deviation from spontaneity also exists in typical (perturbed) Markov processes. In the continuation of such a discussion, let us take as a basis the relation:

$$(p-q)(a-b) \lesseqgtr 0, \quad p > q. \tag{2.20}$$

When $a < b$ the left side is less than zero, when $a = b$ it is equal to zero, and if $a > b$ the left side is greater than zero. Hence, by multiplication we find:

$$pa + qb \lesseqgtr pb + qa.$$

By substituting $a = -\ln p$ and $b = -\ln q$, assuming that we are working with probabilities, we get $p, q \in (0,1)$, and if $p + q = 1$ then it is a binary distribution – with only two outcomes: p the desired outcome occurred and q did not occur. Then:

$$-p\ln p - q\ln q < -p\ln q - q\ln p,$$

and this is the basis of the discrete inequality (2.17) and its corresponding continuous one (2.18). The smaller value, on the left of the inequality, is the Shannon information of the binary distribution, and the larger one on the right is the đisorder". By giving more chances (p) to more information ($-\ln q$), and less (q) to less ($-\ln p$), as in the right-hand side of this relation, we defy the principle of minimalism. This is possible when a system (a living being) has a surplus of options (information) compared to a physical entity (an inanimate being). The success of this analysis is to notice that this surplus does not arise from truth (which cannot lie), but from untruth (by means of vitality).

Hartley information (1.1.2) is the logarithm ($H = \log N$) of the number N of equally likely outcomes. Each of these outcomes has a probability of $p = 1/N$, which is smaller the larger the number of outcomes. We can roughly say the opposite, that we see each probability p as one of some $N = 1/p$ of equal outcomes. Then $H = -\ln p$ is also a type of Hartley information, and Shannon information, in general $S = p_1H_1 + p_2H_2 + p_3H_3 + \dots$, is the mean value of the Hartleys. In the macro world, we work with sets of such ($S_i = -\sum_j p_{ij}\ln p_{ij}$) at once, with sums:

$$Q = a_1S_1 + a_2S_2 + a_3S_3 + \dots = \sum_i a_iS_i = -\sum_i\sum_j a_ip_{ij}\ln p_{ij}.$$

In many natural processes, we use the above models for these approximations. Just as a chemist, and even an astrophysicist, assumes the existence of tiny particles, atoms, and molecules in the objects they study, so too we assume the existence of probabilities here.

The increase in information, $\partial_p(-\ln p) = -1/p$, is opposite to the increase in probability and drastically decreases in the vicinity of impossible events, so we should not be surprised by the absence of "impossible" outcomes in everyday life. If we wanted to (by the power of vitality) realize such, we would need a lot of energy and time, and, as we see, over the centuries of the development of civilization we have succeeded in doing so. In this way, we change inanimate nature and the world around us in a way that it itself will not. The aforementioned partial derivation, information by probability, allows the existence of infinity in uncertainty that real perceptions do not have, but access to them will require ever greater, even immeasurable, efforts.

A special situation is with the probabilities of the squares of the modulus of complex numbers, $p = |z|^2 = z^*z$, where $z = x + iy \in \mathbb{C}$, $z^* = x - iy$, $i^2 = -1$, and $x, y \in \mathbb{R}$. These are regular in quantum mechanics. Then we also have the logarithms of complex numbers:

$$\log z = \log(re^{i\varphi}) = \log r + i\varphi,$$

because the same complex number can also be written in polar coordinates, when $r^2 = x^2 + y^2$ and $\operatorname{tg}\varphi = y/x$, i.e. $x = r\cos\varphi$ and $y = r\sin\varphi$. The angle is usually in radians ($2\pi = 360°$) and the logarithm is then a periodic function with period 2π. This information theory works with all real numbers, including complex numbers, as a measure of the "amount of options".

Where the effect (the product of changed energy and elapsed time) is negative, the information is negative, and when it is complex (an auxiliary result to the real one), the information is also such. The novelty is that we take these "temporary"(auxiliary complex) values more seriouslu here and interpret them as bypass without which there is no reality, just as there is no *completeness of theory* with an insistence on consistency (truths and only truths).

The equivalence of information and action implies the need for an informatic interpretation of the quantum-mechanical scalar product $\langle\psi, \phi\rangle$, where $|\phi\rangle$ and $\langle\psi| = |\psi\rangle^\dagger$ are quantum states. Two vectors $|\psi\rangle$ and $|\phi\rangle$ are considered to determine the same state if and only if $|\psi\rangle = c|\phi\rangle$ for some nonzero complex number $c \in \mathbb{C}$. Observables are given by self-adjoint (Hermitian) operators on a vector space. Not every Hermitian operator corresponds to a physically significant observable, and again, not all physical observables are related to nontrivial self-adjoint (self-adjoint, or Hermitian) operators. For example, mass is a parameter in the Hamiltonian, not a non-trivial (non-zero) operator.

In physics, an *observable* is a physical property or physical quantity that can be measured. In quantum mechanics, observables are Hermitian operators $\hat{A}$, that is, mathematical terms, which when acting on a quantum state $|\psi\rangle$ give any of its eigenvalues λ and a change of state to the eigenstate $\hat{A}|\psi\rangle = \lambda|\psi\rangle$. For example, the position of a particle is an observable wih its eigenvalues ranging from minus

infinity to plus infinity for a free particle. It can take on any of the possible values from a continuous set of values. On the other hand, observables such as spin and angular momentum can have any of only two eigenvalues.

When the operator $\hat{A}$ in quantum mechanics is represented as a matrix, since the choice of the basis of the vector space changes the shape of the matrix, we choose an "eigenbase", that is, the set of eigenvectors $|\psi\rangle$ that are solutions to the eigenequation $\hat{A}|\psi\rangle = \lambda|\psi\rangle$. This is then a diagonal matrix with eigenvalues λ diagonal elements.

By calculating information using a physical action, we relate it to observables, which in quantum physics we treat it with the aforementioned operators and real functions when we observe their sets in the macro world. Then in the relation (2.20) we have $p > q$ and $a > b$, and the left side is greater than zero. Therefore:

$$pa + qb > pb + qa,$$

which means that the information sum of products increases if the corresponding factors in the sum are proportional, when the larger of the elements of the first sequence (p, q) is multiplied by the larger number of the second sequence (a, b), and the smaller is multiplied by the smaller. In this way, we no longer have "information"itself, such as Hartley's (the logarithm of the number of equally likely outcomes) or Shannon's (the mean value of Hartley's), but a scalar product of the state of the physical system, which I call *information of perception*.

This new one is greater the more proportional the participants are, so we could also call it the "information coupling" of the subject and the object of perception, or the possible "amount of communication" of these two states, or the "measure of vitality" if at least one of the participants has it. It measures the strength of the game (the game is a matter of the vital), the mastery of the players, their defiance of inanimate nature or their ability to rule it. Within living beings, or their relationships, for example, in the phenomena of economy, social organization, or wars, the information of perception grows when their subtlety inherent in intelligence and skill is greater.

2.3.7 Recursions

Recursion is defined as a concept or process that depends on a simpler or previous version of itself. We see and use it in a variety of disciplines, from language to logic, including mathematics. Some are very easy.

One of the simpler and often cited examples of recursion is the *factorial* of a natural number, $n! = 1{\cdot}2{\cdot}3\ldots(n-1){\cdot}n$. It is obtained from the recursion $n! = n{\cdot}(n-1)!$. Another, perhaps more famous example is the *Fibonacci sequence*:

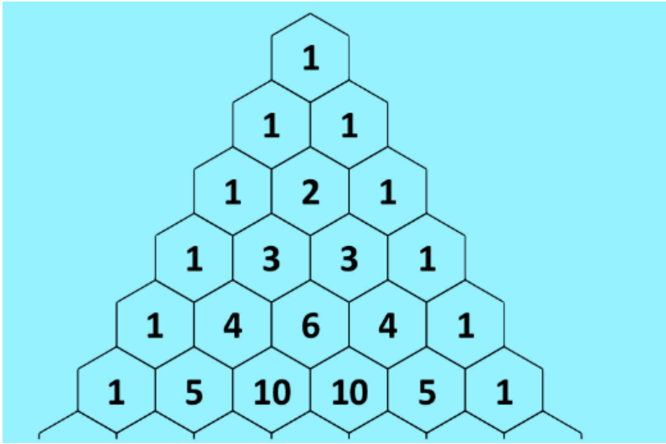
$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \ldots$$

which is defined by the recursion $F(0) = 0$, $F(1) = 1$, $F(n) = F(n-1) + F(n-2)$ for $n > 1$. Computer programming enthusiasts will also demonstrate recursion using the example of the Tower of Hanoi:

Move disk number $n-1$ from the source peg to the auxiliary peg;
Move the nth disk to the destination peg;
Move disk number $n-1$ of the auxiliary to the destination peg.
The powers of the number a can be easily given recursively: $a^0 = 1$, $a^n = a \cdot a^{n-1}$ for $n > 1$. Also the numbers of *Pascal's Triangle*:

$$C(n,0) = C(n,n) = 1, \quad C(n,k) = C(n-1,k-1) + C(n-1,k)$$

and we get a "triangle" of numbers as in the following image (from Pascal's Triangle).



However, there are also more difficult cases to discover using recursion. Here is an easier one of those. If we are looking for the 10th term of the sequence

$$2, 1, 0, 2, 6, 12, 20, ...,$$

then we should first notice that it is built by the recursion $a(n) = n^2 - 3n + 2$, so we find $a(10) = 100 - 30 + 2 = 72$. Now let's move on to the real problems.

Linear recursion of order k with constant coefficients is of the form:

$$a_n = c_{k-1}a_{n-1} + c_{k-1}a_{n-2} + ... + c_0 a_{n-k} + f(n). \tag{2.21}$$

When every $f(n) = 0$ we say that the recursion is homogeneous. It has a simple, explicit general solution expressed by the roots of the characteristic equation:

$$x^k - c_{k-1}x^{k-1} - c_{k-2}x^{k-2} - ... - c_0 = 0.$$

Let $P(x)$ be the polynomial on the left-hand side of this equation. It has a root, say, r of multiplicity m by which it factors $P(x) = (x-r)^m Q(x)$ with some polynomial $Q(x)$ of degree m lower than its own. For example, $P(x) = (x+1)^2(x-5)$ is a polynomial with two roots, $r_1 = -1$ of multiplicity 2 and $r_2 = 5$ of multiplicity 1.

Theorem 78. *The general solution of linear recursion of order k with constant coefficients (2.21) is*

$$a_n = P_1(n)r_1^n + P_2(n)r_2^n + ... + P_j(n)r_j^n$$

where $r_1, r_2, ..., r_j$ are the roots of multiplicity of order $m_1, m_2, ..., m_j$ of the characteristic equation, and P_i is an arbitrary polynomial of degree $m_i - 1$ for each $i = 1, ..., j$.

The proof of this theorem follows directly from the inclusion of the solution in the recursion, and I will not present it here, since it is a well-known and routine operation. We prefer to see how its solution is used.

Example 79. *Let's find the general solution of the recursion*

$$a_n = 3a_{n-1} + 9a_{n-2} + 5a_{n-3}$$

and then the particular solution for $a_0 = 0$*,* $a_1 = 1$ *and* $a_2 = 3$*.*

Solution. This is a homogeneous linear recursion of the third order ($k = 3$) of constant coefficients, and the characteristic equation is:

$$x^3 - 3x^2 - 9x - 5 = 0.$$

The left-hand side is a polynomial $P(x) = (x+1)^2(x-5)$ with root $r_1 = -1$ of multiplicity $m_1 = 2$ and root $r_2 = 5$ of multiplicity $m_2 = 1$. Therefore, the general solution of the recursion is

$$a_n = (\alpha n + \beta)(-1)^n + \gamma 5^n,$$

where α, β, γ are arbitrary constants. Finding a given particular solution also requires solving the corresponding system of linear equations:

$$\begin{cases} \beta + \gamma = 0 \\ -\alpha - \beta + 5\gamma = 1 \\ 2\alpha + \beta + 25\gamma = 3 \end{cases}$$

Hence, $\alpha = -1/6$, $\beta = -5/36$ and $\gamma = 5/36$. Therefore:

$$a_n = (-1)^n(-\frac{1}{6}n - \frac{5}{36}) + 5^n \cdot \frac{5}{36}$$

is the required particular solution. □

An essential feature that determines the past is the taking of information from the present. It thus deprives future events of some of the "quantity of possibilities" and thus partly manages them, increases the certainty of the present, and directs it. It sounds as if we are talking about recursions. Also, more than a metaphor, recursion is a story about transmission channels with the accumulation of interference and especially such stochastic matrices in Markov chains. That is one of the reasons for this "diversion" off topic.

Example 80. *Given a stochastic matrix*

$$A = \begin{pmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{pmatrix}.$$

Treating as a recursion, find A^n*.*

Solution. We are looking for a matrix of the form:

$$A^n = \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix}, \quad n = 0, 1, 2, ...,$$

$$A^{n+1} = \begin{pmatrix} a_{n+1} & b_{n+1} \\ c_{n+1} & d_{n+1} \end{pmatrix} = \begin{pmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{pmatrix}^n \begin{pmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{pmatrix} =$$

$$= \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix} \begin{pmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{pmatrix} = \begin{pmatrix} 0.8a_n + 0.2b_n & 0.4a_n + 0.6b_n \\ 0.8c_n + 0.2d_n & 0.4c_n + 0.6d_n \end{pmatrix}.$$

Thus we find the recursions, the differential equations:

$$\begin{cases} a_{n+1} = 0.8a_n + 0.2b_n \\ b_{n+1} = 0.4a_n + 0.6b_n \end{cases} \qquad \begin{cases} c_{n+1} = 0.8c_n + 0.2d_n \\ d_{n+1} = 0.4c_n + 0.6d_n \end{cases}$$

From the first two, by elimination, we find:

$$a_{n+2} = 0.8a_{n+1} + 0.2(0.4a_n + 0.6b_n) =$$

$$= 0.8a_{n+1} + 0.08a_n + 0.6(0.2b_n) = 0.8a_{n+1} + 0.08a_n + 0.6(a_{n+1} - 0.8a_n),$$

$$a_{n+2} = 1.4a_{n+1} - 0.4a_n.$$

This is a recursion with the characteristic equation $x^2 - 1.4x + 0.4 = 0$ with roots $r_1 = 1$ and $r_2 = 0.4$ both single. The general solution and another analogue are:

$$a_n = \alpha + \beta \cdot 0.4^n, \quad c_n = \gamma + \delta \cdot 0.4^n.$$

How is it:

$$A^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad A^1 = \begin{pmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{pmatrix}, \quad A^2 = \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix}, \dots$$

we calculate $\alpha = 2/3$, $\beta = 1/3$, $\gamma = 1/3$, $\delta = -1/3$, then the other two unknowns:

$$A^n = \frac{1}{3} \begin{pmatrix} 2 + 0.4^n & 2 - 2 \cdot 0.4^n \\ 1 - 0.4^n & 1 + 2 \cdot 0.4^n \end{pmatrix}.$$

For each n, A^n is a stochastic matrix. □

We would get the same by searching for the diagonal matrix (example 50).

Example 81. *Diagonalize A of the previous, 80th example, and find A^n.*

Solution. We solve the eigenvalue $\det(A - \lambda I) = 0$:

$$\begin{vmatrix} 0.8 - \lambda & 0.4 \\ 0.2 & 0.6 - \lambda \end{vmatrix} = 0,$$

$$\lambda^2 - 1.4\lambda + 0.4 = 0.$$

It is not surprising that this is also the characteristic equation of the previous recursion, now with eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 0.4$. These different values guarantee independent corresponding ($Av = \lambda v$) eigenvectors. We calculate them too:

$$\begin{pmatrix} 0.8-\lambda & 0.4 \\ 0.2 & 0.6-\lambda \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 0,$$

$$\begin{cases} (0.8-\lambda)x + 0.4y = 0 \\ 0.2x + (0.6-\lambda)y = 0 \end{cases}$$

from which, by alternating the first and second lambdas, we find:

$$v_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

The auxiliary matrix $P[v_1, v_2]$ and its inverse P^{-1} give the diagonal ($D = P^{-1}AP$):

$$D = \frac{1}{3}\begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0.4 \end{pmatrix}.$$

This was a check. Now we write $A = PDP^{-1}$ and raise it to a power:

$$A^n = (PDP^{-1})^n = PD^nP^{-1} =$$

$$= \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0.4^n \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}\frac{1}{3} =$$

$$= \begin{pmatrix} 2 & 0.4^n \\ 1 & -0.4^n \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}\frac{1}{3}$$

$$= \frac{1}{3}\begin{pmatrix} 2+0.4^n & 2-2\cdot 0.4^n \\ 1-0.4^n & 1+2\cdot 0.4^n \end{pmatrix}.$$

which is equal to the previous solution. □

When you notice the wealth of methods in solving these problems, I hope you will be able to agree with me more easily about the wealth of interpretations of the solutions themselves. To begin with, it would be enough to see information as observable. This is another reason for drawing recursions in the topic of information coupling, I remind you, in addition to the aforementioned metaphors with the melting of the present into the past and the connection of information transfer through Markov chains.

2.3.8 Periodicity

Determinant is the product of the eigenvalues of the matrix (operator), and stohastic matrices havve eigenvalues of modulus not greater than one. On the other hand, as long as the determinant is nonzero, the connected links have some kind of transmission. In other words, the Markov chain does not tend to a black box when the determinant of the generator of the modulus chain is one.

Let S be a stochastic matrix that after $\mu \in \mathbb{N}$ transitions becomes the matrix $M = S^\mu$ of the Markov chain. Then, the following equality holds for the determinants:

$$\det M = \det S^\mu = (\det S)^\mu,$$

so if $\det S = 1$, then $\det M = 1$ no matter how long the chain is, and there is no black box. The closer $\det S$ is to the unit norm, the slower M converges to that *steady state*, and the more links ($\mu \to \infty$) are needed to make $\det M \approx 0$ with sufficient accuracy. At the other extreme, if $\det S = 0$, then everything is over at the very beginning; there is no data transfer even in the first step.

In the presentation of "Information Theory I" [13] (03. Q Channel Matrix), I considered *quasi-stochastic* matrices, say like this one of type 2×2, designed for single use:

$$Q\vec{v} = \begin{pmatrix} \cos\beta_1 \cos\alpha_1 & \cos\beta_1 \cos\alpha_2 \\ \cos\beta_2 \cos\alpha_1 & \cos\beta_2 \cos\alpha_2 \end{pmatrix} \begin{pmatrix} \cos\alpha_1 \\ \cos\alpha_2 \end{pmatrix} = \begin{pmatrix} \cos\beta_1 \\ \cos\beta_2 \end{pmatrix} = \vec{w},$$

because $\cos^2\alpha_1 + \cos^2\alpha_2 = 1$. It resembles quantum mechanics because the squares of the elements are probabilities. However, it does not have an inverse matrix ($\det Q = 0$) and does not describe "reality" in the sense that it does not imitate the conservation law. In doing so, its transposed matrix performs the inverse transformation $Q^\top \vec{w} \to \vec{v}$, acting as if it were its inverse, although it is not. I mention it as a curiosity, as well as the next one.

$$\det \begin{pmatrix} 1 & a & a \\ 0 & b & b \\ 0 & c & c \end{pmatrix} = 0.$$

The determinant of a stochastic matrix is a number from 0 to 1 inclusive (theorem 69). It is equal to 1 if the matrix is permutation, where the determinant itself is equal to 1 for an even permutation and -1 for an odd permutation. Conversely, when the determinant is zero, then its columns are linearly dependent and its matrix is not reversible. Reffered to as a state transition matrix, it irreversibly loses some of the information it has passed after just one step.

Example 82. *Coefficients of the stochastic matrix $S = (s_{ij})$ less than one narrow the range of possible probabilities in the output distribution.*

Proof. Let m_j and M_j be the minimum and maximum values of the j-th column of the given channel matrix $S = (s_{ij})$, with indices $i, j = 1, 2, \dots, n$. From $q = Sp$ it follows:

$$m_j = \min_{1 \le i \le n}\{s_{ij}\}, \quad M_j = \max_{1 \le i \le n}\{s_{ij}\}$$

$$m_j = \sum_{i=1}^{n} m_j p_i \le s_{ij} p_i \le \sum_{i=1}^{n} M_j p_i = M_j$$

$$m_j \le \sum_{1 \le i \le n} s_{ij} p_i \le M_j$$

$$m_j \le q_j \le M_j.$$

These inequalities do not depend on the input distribution $p = (p_1, p_2, ..., p_n)$ nor on the output $q = Sp = (q_1, q_2, ..., q_n)$, but only on the minimum and maximum of the j-th column of the channel matrix. □

The example is from the book [6] written before the war in Bosnia and Herzegovina (1991-1995) and printed afterwards. Here it is in support of the claim that the determinant of a stochastic matrix can be 1 only if it is a permutation matrix. In the same book there is also a štronger", direct proof of such a position.

Theorem 83. *A stochastic matrix $A = (a_{ij})$ of order n has an inverse stochastic matrix $B = (b_{ij})$ if and only if it is obtained by permuting the columns (rows) of the identity matrix $I = (\delta_{ij})$, where $\delta_{ii} = 1$ and $\delta_{ij} = 0$ when $i \ne j$. The inverse matrix (B) is then the transpose of the given (A), so $B = A^\top = A^{-1}$ and $AA^\top = A^\top A = I$.*

Proof. Assume that there exists B such that $AB = I$, i.e.

$$\sum_{k=1}^{n} a_{ik} b_{kj} = \delta_{ij}.$$

By assumption, matrices A and B are stochastic, so all their coefficients are non-negative, and each is a unique ($k = 1, 2, ..., n$) sum $a_{ik} b_{kj} = 0$ for $i \ne j$. Since B is stochastic, it will have at least one nonzero coefficient in each column. Let this be the j-th column and $b_{kj} \ne 0$, so we divide the above equality by this coefficient. We get

$$a_{ik} = \begin{cases} 1, & j \ne i \\ 0, & j = i \end{cases}$$

where $a_{kj} = 1$, because the matrix A is stochastic. Therefore, each column of this matrix has one unit, and all other coefficients are zero.

The matrix A has exactly n of these units. If its type had exactly all coefficients zero, then the product AB would not give a unit matrix. Therefore, each type of matrix A has one unit. Therefore, it is obtained by permutations of the columns (type) of the unit matrix.

Finally, starting from the above equality, for $i = j$ we obtain $\sum_m a_{im} b_{mj} = 1$, and $a_{ik} b_{ki} = 1$ for each individual k, because all other sums (for $m \ne k$) are zero. Hence the equality a_{ik} gives $b_{ki} = c$, where $c \in \{0, 1\}$, which is by definition a transpose. We have thus also proved that the matrix B is the transpose of A and that both are obtained by permutations of the columns of the identity matrix. □

That this theorem is valid only for square (stochastic) matrices is demonstrated by the product of non-square matrices (for all a, b):

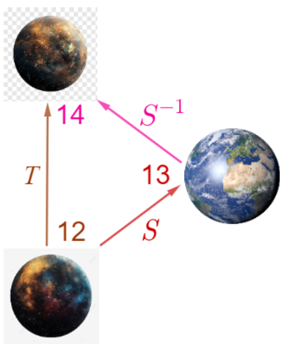
$$\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 1\\ a & b\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$

which can be transposed (when $a = b = 0$) and the result is the identity matrix. There are two (2!) second-order stochastic matrices that are inverse to themselves:

$$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \quad \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}.$$

They are also symmetric ($p_{ij} = p_{ji}$). These are the only second-order square matrices that have an inverse stochastic matrix. There are six ($3! = 1 \cdot 2 \cdot 3$) third-order stochastic matrices with an inverse stochastic matrix, which are permutations of the columns of the identity matrix, two of which are not symmetric. There are $n! = 1 \cdot 2 \cdot 3 ... n$ n-order stochastic matrices that have an inverse stochastic matrix, and all of them are permutations of the columns of the identity matrix.

However, the nature of information would not be what it is without its multiplicity and the ability to integrate these abstract theories almost everywhere around us. For example, celestial bodies (planets, constellations) become smaller the further they are from us. As in the case of a typical (with interference) Markov chain, images with increasing distance from the object have less and less of the object. Suffocated by the "noise"of the environment, they tend to become a "black box"with only the environment coming out.



On the other hand, let's say a particle of light traveling from here to there takes forever, so we have a coupled isometry (lossless transmission) with losses during transmission. The determinant of the isometry is 1, because by chaining copies of the power operator, the determinant of the total transmission would either converge to zero (if its modulus is less than 1), or diverge to infinity (if its modulus is greater than 1). The determinant of a typical stochastic matrix is less than one. In the figure on the left, the transfer S of an image of a planet from the 12th billion-year-old universe to us in the 13th and the “inverse” transfer S^{-1} to the 14th. In fact, we don't need to look at the stars to notice this absurdity; it also happens with objects in our immediate environment, because light takes time to travel.

So, the sequence of images of an increasingly distant object is not reversible. First, theoretically, because the determinant of the noisy channel is not 1, and then practically, because going there does not mean finding what was sent. Therefore, the following matrices are inverse ($S \cdot S^{-1} = I$):

$$\begin{pmatrix}0.6 & 0.4\\ 0.4 & 0.6\end{pmatrix} \cdot \frac{1}{0.20}\begin{pmatrix}0.6 & -0.4\\ -0.4 & 0.6\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$

and only the first (S) is stochastic. The first is even doubly stochastic, and the second is not at all. There is no information transfer by the inverse matrix (S^{-1}) from "today"to "yesterday", although there is a way to decipher "yesterday"from "today". The determinant of the first matrix is $\det S = 0.20$ and its inverse $\det S^{-1} = 1$, which goes with the inverse of the inverse being stochastic while the inverse of the stochastic is not stochastic.

These conclusions do not change even when the double stochastic matrix (S) is replaced by an ordinary stochastic matrix ($A \cdot A^{-1} = I$) as in the following multiplication:

$$\begin{pmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{pmatrix} \cdot \frac{1}{0.30} \begin{pmatrix} 0.6 & -0.4 \\ -0.3 & 0.7 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

However, we will notice that the matrix that is "more diagonal", that has larger diagonal elements, has a larger determinant (here $\det S \to \det A$, i.e. $0.20 \to 0.30$). Here is a reinforcement of the previous theorem in this sense.

Theorem 84. *Let $\vec{x} = (\xi_1, \xi_2, ..., \xi_n)$ and $\vec{y} = (\eta_1, \eta_2, ..., \eta_n)$ be vectors of two arbitrary distributions, but such that $\vec{y} = S\vec{x}$, where $S = (s_{ij})$ is a double stochastic matrix with eigenvector ($S\vec{v} = \vec{v}$) $\vec{v} = (\nu_1, \nu_2, ..., \nu_n)$. Then*

$$\max_{1\le k\le n} |\nu_k - \eta_k| \le \max_{1\le j\le n} |\nu_j - \xi_j|$$

where the equality holds if and only if from $\nu_i - \xi_i < \max_j |\nu_j - \xi_j|$ it follows that $s_{ij} = 0$ for each $i = 1, 2, ..., n$.

Proof. By subtraction we obtain $\vec{v} - \vec{y} = S(\vec{v} - \vec{x})$. In this system we observe the m-th equation where the absolute difference $|\nu_m - \eta_m|$ is maximal, i.e. the equation

$$\nu_m - \eta_m = \sum_k (\nu_k - \eta_k) s_{kj}.$$

Since

$$\sum_k s_{kj} = 1,$$

we further obtain:

$$\max_k |\nu_k - \eta_k| = |\nu_m - \eta_m| \le \sum_k s_{ik} |\nu_k - \xi_k| \le \max_k |\nu_k - \eta_k|$$

which was to be proven. □

Everything said here about channel matrices, whose coefficients are real numbers between 0 and 1, is easily transferred to complex matrices. The theorems then apply to the modulus of the coefficients, or rather the squares of the modulus (of quantum mechanics). What we could get by switching to complex numbers speaks of the "principle of *minimalism* information" in the language of matrix algebra. As vectors (states) and their linear operators (processes) converge towards distributions of "quieter" scattering, smaller dispersions of outcome probabilities, we will say that *adapt*.

When the message adaptation is monotonic, advancing each state through the links of a Markov chain (such as a Cauchy sequence) to an eigenstate of the channel generator matrix at each step, if there is no backsliding, then the resulting output converges to the eigenstate and the overall channel becomes a black box. But if this is not the case, when the determinant of the channel matrix is 1, then the message travels *periodically*, changing, taking previous states over and over again.

Namely, due to the finiteness of perceptions, no matter how large the number of their permutations, the exhaustion of all possibilities will occur. However, there is no ordering of all options when the principle of minimalism operates in diversity. Old news repeated is no longer the same news and will transform into one of its closer forms. It is like a photon that oscillates between electric and magnetic forms as it travels, or in general like *waves of probability* (a term of quantum mechanics) that, circling, goes through more likely forms, and they are always current in the multitude of different ones.

For example, consider the degrees of the stochastic matrix S, permutations:

$$S = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad S^2 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad S^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

where $S^3 = I$, and then again $S^4 = S$, $S^5 = S^2$ and so on periodically. This is a stochastic matrix of the third order ($n = 3$), but analogously it happens with much larger matrices ($n \in \mathbb{N}$) that are closer to reality. Then such cycles can be much longer, but not infinite. Let us not confuse various rotations, as photon states, with changes in relation to the environment that do not take previous forms. No subject has the same information of perception of the wider coupling.

2.4 Continuum

Continuum is a compact set that cannot be separated into two sets, neither of which contains the limit point of the other. Such a set is the set of real numbers $\mathbb{R}$, which includes both rational and irrational numbers. Rational numbers are those numbers from the set $\mathbb{Q}$ that can be expressed as an integer $\mathbb{Z} = \{0, \pm1, \pm2, \dots\}$, or the quotient of an integer with a natural number $\mathbb{N} = \{1, 2, 3, \dots\}$. An irrational number is one that can be written as an infinite sequence of decimals without periods of consecutive digits that are constantly repeated and cannot be expressed as the quotient of two integers.

If each element $x \in X$ corresponds in some way to a certain element $y \in Y$, we say that the set X is mapped into the set Y. This x is the original, and y is its image. If we denote this mapping by f, we also write the image $f(x)$. We also write $x \to f(x)$, or $f : X \to Y$. Functional analysis is a part of mathematics (in teaching, it is considered its more difficult part) that studies and proves the properties of functions so that we can freely use its results without fear of being deceived by some untruth. Here are some of its introductory points.

Example 85. *Given a function $f : X \to Y$ and subsets $X_1, X_2 \subset X$, then:*

1. $X_1 \subseteq X_2 \Rightarrow f(X_1) \subseteq f(X_2)$;
2. $f(X_1 \cup X_2) = f(X_1) \cup f(X_2)$;
3. $f(X_1 \cap X_2) \subseteq f(X_1) \cap f(X_2)$.

If f takes the same value on the differences of the sets $X_1 \setminus X_2$ and $X_2 \setminus X_1$, but not on the intersection $X_1 \cap X_2$, then in 3. the strict inclusion $f(X_1 \cap X_2) \subset f(X_1) \cap f(X_2)$ holds.

Subset $A \subset B$ means the same as the logical $(\forall x)\ x \in A \Rightarrow x \in B$. The union $A \cup B$ is equivalent to $x \in A \vee x \in B$. The intersection of $A \cap B$ with $x \in A \wedge x \in B$, and the difference $A \setminus B$ is the value of $x \in A \wedge x \notin B$. The reciprocal, *inverse image* of an element $y \in Y$ is the set of all elements $x \in X$ for which $f(x) = y$, and we denote it by $f^{-1}(y)$. The reciprocal image can also be the empty set, whenever $y \notin f(X)$.

Example 86. *For the inverse (reciprocal) image:*

1. $Y_1 \subseteq Y_2 \subseteq Y \Rightarrow f^{-1}(Y_1) \subseteq f^{-1}(Y_2)$;
2. $f^{-1}(Y_1 \cup Y_2) = f^{-1}(Y_1) \cup f^{-1}(Y_2)$;
3. $f^{-1}(Y_1 \cap Y_2) = f^{-1}(Y_1) \cap f^{-1}(Y_2)$;
4. $f^{-1}(Y_1^c) = (f^{-1}(Y_1))^c$.

A^c is a *complementary to the set* A, i.e. $x \in A \iff x \notin A^c$. According to this, 4. originals whose copies are the complement of Y_1 are the same as the complements of the originals copied into Y_1. According to the 3rd of the same examples, the originals of the intersections of the copied ones are equal to the intersection of the originals that give those copies. Conversely, in the previous example 3, it is said that the image of the intersection of the original is a subset of the intersection of the images. Namely, the copies from both X_1 and X_2 can be the same even then in the copy of the intersection, but the intersection of the copies involves more originals. The difference between these 3 items of the two examples speaks of a possible reduction or impoverishment of the images in relation to the originals.

2.4.1 Banach space

In mathematics, its functional analysis, Banach space is a *complete* normed vector space. It has a metric that gives the lengths of vectors and distances between points (vectors), while completeness, we also say compactness, means that every Cauchy sequence of its points converges, goes to the limit which is itself a point of that space. For a more precise definition of a "vector space", see here in the section on vector multiplication.

Elements of a vector space are called vectors. Examples of vectors are numbers, sequences, matrices, and polynomials, solutions of differential equations, or wave

functions of quantum mechanics, and linear operators. The list of possible vector spaces would be very long, and therefore the study of the consequences of their four axioms is more important. Not every vector space has to have a norm, but it can have different ones.

The usual vector norms of a vector $x = (\xi_1, \xi_2, \dots, \xi_n)$ are p-norms:

$$\|x\|_p = \left(\sum_{k=1}^{n} |\xi_k|^p\right)^{1/p}, \quad p \ge 1.$$

The most common of these are $p = 1$ (Manhattan Norm) when L_1 is the label space, or $p = 2$ (Euclidean Norm) of the label space L_2, or $p \to \infty$ (Max Norm) whose space is then L_∞. The latter comes from the limit value:

$$\|x\|_\infty = \lim_{p\to\infty} \|x\|_p = \max_{1\le k\le n}\{|\xi_k|\}.$$

Each satisfies the four norm axioms, but the norm $\|x\|_p$ is a decreasing function of the parameter $p \ge 1$. For example, for a binary sequence $x = (3,4)$ we have $\|x\|_1 = |3| + |4| = 7$, and $\|x\|_2 = \sqrt{|3|^2 + |4|^2} = 5$ and $\|x\|_\infty = \max\{3,4\} = 4$, so $\|x\|_\infty \le \|x\|_2 \le \|x\|_1$.

When $n \to \infty$ we denote the spaces of these norms by ℓ_p. When the scalars are real numbers, we emphasize this by denoting the spaces by $\mathbb{R}_p^n$, or $\mathbb{R}_p^\infty$. If the scalars are complex, by $\mathbb{C}_p^n$ or $\mathbb{C}_p^\infty$. The space of continuous functions $f(t)$ on a given interval (a,b) has the label and norm, respectively:

$$L_p(a,b), \quad \|f\| = \left(\int_a^b |f(t)|^p \, dt\right)^{1/p},$$

and in the case of $p \to \infty$ the labels of the space and the norm are:

$$C[a,b], \quad \|f\| = \max_{a\le t\le b} |f(t)|.$$

When $n \to \infty$ and the interval can be open, this space is $M(a,b)$, and the norm $\sup|f(t)|$ is the smallest of the larger values. Based on the given ones, it is possible to create various combined norms that we also use, for example:

$$\|x\| = \sup_{0\le t\le 1} \left|\sum_{k=1}^{n} \xi_k t^{k-1}\right|, \quad \|x\| = \int_0^1 \left|\sum_{k=1}^{n} \xi_k t^{k-1}\right| w(t)\, dt,$$

where $w(t)$ is a continuous function of some positive weights (weight).

Quantum mechanics uses the bracket notation and the $p = 2$ norm for the scalar product of quantum states and vectors:

$$\langle x, y\rangle = \sum_k \xi_k \eta_k^*, \quad \langle f, g\rangle = \int f(t) g^*(t)\, dt,$$

or possibly:

$$\langle x, y\rangle_w = \sum_k \xi_k \eta_k^* w_k, \quad \langle f, g\rangle_w = \int f(t) g^*(t) w(t)\, dt,$$

depending on whether the vectors are sequences or functions, if with weight w a sequence or function. Consistently, the norm of its quantum space is written as $\|x\| = \sqrt{\langle x, x\rangle}$. Scalars in quantum physics are complex numbers, $\Phi = \mathbb{C}$, and vectors are sequences or (continuous) wave functions. Operators and their Hermitian matrices A and B can also form a scalar product $\langle A, B\rangle = \mathrm{Tr}(AB^\dagger)$. The aforementioned operator trace is otherwise the sum of the eigenvalues of the operator (matrix) where each repeats with its multiplicity.

When a vector space X is supplied with a norm $\|x\|$, then $d(x, y) = \|x - y\|$ is a metric on X. The set X supplied with a metric d is called a metric space (it can also be a vector space), its elements are called points (vectors), and $d(x, y)$ is the distance between points x and y. The metric introduced using the norm of a vector space has some additional properties that metric spaces themselves do not necessarily have. The translation invariance $d(x + v, y + v) = d(x, y)$ and the scaling $d(\lambda x, \lambda y) = |\lambda| d(x, y)$ are such. In such a metric space, the norm can be introduced with $\|x\| = d(x, 0)$.

It is also useful to know the following few terms *topologies* (geometry without a defined metric) that are used in functional analysis and beyond. *Open set* is a generalization of the open interval of the real line, which contains all the interior points of a sphere without a surface sphere. *Countable* are infinite sets of natural numbers $\mathbb{N} = \{1, 2, 3, ...\}$, integers $\mathbb{Z} = \{..., -2, -1, 0, 1, 2, ...\}$, or rational numbers $\mathbb{Q} = \{\frac{m}{n} : m \in \mathbb{Z} \land n \in \mathbb{N}\}$. Their infinity is denoted by $\aleph_0$ (read as "aleph-zero). Uncountably infinite are the real numbers $\mathbb{R}$ whose set, or quantity, is called the *continuum*, denoted by $\mathfrak{c}$.

A space is *Separable* if it contains a countable, dense subset; that is, there is a sequence of elements of the space such that every nonempty open subset contains at least one element of that sequence. An example of a separable space is the real axis with a rational ($\mathbb{Q}$) set of points. A space that is not separable is a discrete set, say $\mathbb{Z}$ on the real axis.

Adherent point (connection point, closure or contact point) of a subset A of a topological or metric space X is a point $x \in X$ whose every neighborhood, or every open neighborhood, contains at least one point from A. A point $x \in X$ is an adherent of A if it is in the *closure* of the set A, i.e. in the set obtained by adding A to all the limit values of its sequences. *Compactness* follows from the properties of a subset of Euclidean space that is closed (has all the limit points of its sequences) and bounded (all the points of which lie within a given fixed distance).

Unlike the discrete energy levels that only an electron trapped in an atom can have, a free particle is not limited to taking on, theoretically speaking, countless possibilities. The first demonstrates the coupling and limitation of perceptual information, and the second the infinity of their environments. In this theory, we hold to the "reality"of both and therefore consider each of them equally.

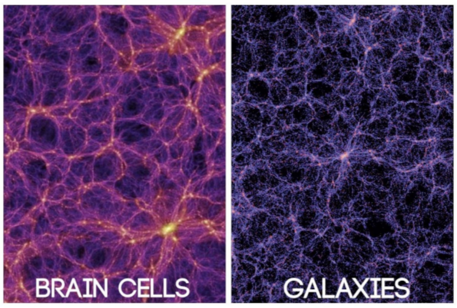
I remind you that in this information theory, “real” is any of the following: 1. that which can be observed; 2. that for which some form of conservation law holds; 3. that which is true. This implies a layering of the concept of “real” and the acceptance of infinity as real. Therefore, it is not expected that in the further development of such ideas we will remain within the framework of classical applications

of mathematics.

2.4.2 Uniqueness

I have had rare occasions to underline "the truth is everyone's, and a lie is always someone's,"only to insert myself into someone else's context of some more lively but not particularly cognitive discussion, simply to emphasize the difference between Skolem's and Riesz's theorems. Then I would think of the infinite forms of correct theories in nature around us, in contrast to the uniqueness of the subjects of observation.

The image on the right shows the discovery of similarities in the operation of two systems of completely different scales – the network of neuronal cells of the human brain and the cosmic network of galaxies (Foglets). These phenomena are not just "déjà vu" (French: already seen) but a "reality" that arises from the first mentioned Löwenheim-Skolem theorem, which indicates the unlimited occurrence of parts of abstract truths in otherwise unrepeatable, very concrete physical objects, and on the other hand from (my) new teaching about the reality of truth.



Namely, as every truth is permanent, so untruth is temporary. The first is reality and the second is fiction. Both exist, but the first can be observed, and the second is intuited through *vitality* (the excess of options we have in relation to the physical substance of which we are composed), the lies associated with us that increase the dead substance's powers of choice and prolong its own survival in that world. Truth participates in physical interactions in such a way that when we prove that it is not present in a given experiment, then the desired experience is not present either. A lie is a manifestation only of our fictions thanks to vitality.

It would be difficult to put all this together if it weren't for the ideas from quantum mechanics that physical states are interpretations of vectors. At the micro level, rare and small phenomena are observable, as well as elementary particles, but we won't say that large physical bodies don't consist of such (quanta, atoms, molecules) just because they are tiny and we don't pay attention to them. Their constancy (reality) extends through the scale of magnitudes, not only through space-time translation. The same is true with vectors and quantum states, I believe.

Well, the states of physical objects are interpretations of vectors. They can have a very large number ($n \in \mathbb{N}$) of independent components and such a large base of vector space whose representation is reality that it is not possible, or makes no sense, to calculate them all, and yet by studying them we learn something that we could not otherwise. We start from the idea that the information of perception is some functional and scalar products, and in the following example we work only

with $n = 3$ (independent) observables. The example will explain Riesz's theorem.

Example 87. *The functional $f : X \to \Phi$ maps the following vectors $x_1 = (3, -2, 1)$, $x_2 = (2, -1, 0)$ and $x_3 = (1, 0, 1)$ to the numbers 2, 0 and 4 respectively. To what number does this functional map the vector $x_0 = (4, 5, 6)$?*

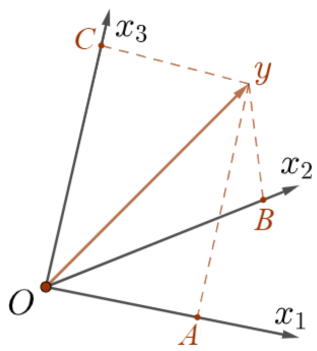
Solution. We solve the system of equations:

$$\begin{cases} 3\eta_1 - 2\eta_2 + \eta_3 = 2, \\ 2\eta_1 - \eta_2 = 0, \\ \eta_1 + \eta_2 = 4, \end{cases}$$

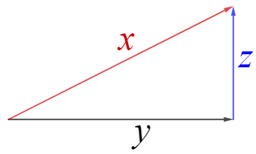
and find $\eta_1 = 1$, $\eta_2 = 2$ and $\eta_3 = 3$. Therefore, the functional is $f(x) = \langle x, y \rangle$, where $x = (\xi_1, \xi_2, \xi_3)$ is an arbitrary vector (which f maps to a scalar), and $y = (1, 2, 3)$ is a fixed vector that defines this functional. So, this functional maps the given vector x_0 into the number $f(x_0) = 4 \cdot 1 + 5 \cdot 2 + 6 \cdot 3 = 32$. $\square$

This is what this concept of uniqueness is about. In the example, we interpret the found vector y as a "subject" whose total observations of the surrounding states, the vector $x \in X$, define the functional f. The conditions for the uniqueness of the subject with the given $n = 3$ mappings speak of an n-dim space of independent observables. If the number of independent observables were only one larger, or if the three vectors x_1, x_2, and x_3 were not linearly independent, the system of equations in the example would give an infinite number of solutions. This is precisely the mysterious difference between the infinity of the general and the uniqueness of the particular.

Namely, in contrast to the limitations of personal perceptions, truths are always so universal that every subjective system of perception is incomplete with them. In the words of Gödel's incompleteness, truths are too consistent and therefore phallic. Or, on the other hand, for our perceptions, the rest of "reality"is in such darkness that we can live as if it does not exist. This dictates Borel–Cantelli lemma, that in a convergent order with probability 1 only finitely many outcomes can be realized, and not, say, the non-existence of what we do not see.



In the figure on the left we see that the projections at points A, B, C onto linearly independent vectors x_1, x_2, x_3 in order uniquely determine the vector y of 3-dimensional space. If the given x-vectors were linearly dependent, say that all three lie in the plane ABO, then on the plane perpendicular to it, which would cut it along the line AB, there would be an infinite number of (vertices) of vectors y, all of which would solve the problem. There is no uniqueness when some of the observables that build its reality are not available to the subject. On the other hand, for general truths, that shirt is tight; however, any larger one is tight again. I will explain this with a few well-known theorems of linear algebra. We are working in L_2 space.



Orthogonal decomposition is the decomposition of a given vector (x) into the sum of two mutually perpendicular vectors (y and z), such that $x = \lambda y + z$ and $\langle z, y\rangle = 0$, as in the following figure on the right. It is always $x = \lambda y + (x - \lambda y)$, and we easily find $0 = \langle x - \lambda y, y\rangle = \langle x, y\rangle - \lambda \|y\|^2$, then $\lambda = \langle x, y\rangle / \|y\|^2$, and with $z = x - \lambda y$ it will be $x = \lambda y + z$ and $\langle z, y\rangle = 0$.

For $e_1, e_2, \dots, e_n \in X$ the orthonormal list of vectors will be:

$$\|a_1e_1 + a_2e_2 + \dots + a_ne_n\|^2 = |a_1|^2 + |a_2|^2 + \dots + |a_n|^2,$$

for all $a_1, a_2, \dots, a_n \in \Phi$. This follows from the previously used decomposition (Pythagoras' theorem). Because the given vectors (e_k) form a basis, there are scalars (a_k) for an arbitrary vector

$$x = a_1e_1 + a_2e_2 + \dots + a_ne_n,$$

and since the basis is orthonormal, taking the inner product with e_j on both sides of the equality we obtain $\langle x, e_j\rangle = a_j$, and then:

$$x = \langle x, e_1\rangle e_1 + \langle x, e_2\rangle e_2 + \dots + \langle x, e_n\rangle e_n, \quad \|x\|^2 = |\langle x, e_1\rangle|^2 + |\langle x, e_2\rangle|^2 + \dots + |\langle x, e_n\rangle|^2.$$

This convenience of orthonormal bases is generally used by the Gram-Schmidt procedure, which transforms a set of linearly independent vectors into a list of orthonormal vectors in the same space by a series of operations. Here we will use it to prove Riesz's theorem (Frigyes Riesz, 1880 - 1956).

Theorem 88. *If f is a function on a finite-dimensional space X, then there exists a unique vector $y \in X$ such that $f(x) = \langle x, y\rangle$ for every $x \in X$.*

Proof. First, let us show that there exists a vector $y \in X$ such that $f(x) = \langle x, y\rangle$ for every $x \in X$. Let $e_1, \dots, e_n$ be an orthonormal basis in X. Then, continuing from the previous:

$$f(x) = f(\langle x, e_1\rangle e_1, \dots, \langle x, e_n\rangle e_n) = \langle x, e_1\rangle f(e_1) + \dots + \langle x, e_n\rangle f(e_n) =$$

$$= \langle x, f^*(e_1)e_1 + \dots + f^*(e_n)e_n\rangle, \quad \forall x \in X.$$

Therefore, there exists a vector $y = f^*(e_1)e_1 + \dots + f^*(e_n)e_n$ such that $f(x) = \langle x, y\rangle$ for every vector $x \in X$.

Next, let us prove that only one vector $y \in X$ has the desired property. Suppose that for two vectors y_1 and y_2 from X we have $f(x) = \langle x, y_1\rangle = \langle x, y_2\rangle$ for each $x \in X$. Then $0 = \langle x, y_1\rangle - \langle x, y_2\rangle = \langle x, y_1 - y_2\rangle$ for each $x \in X$, including $x = y_1 - y_2$, which means $\|y_1 - y_2\|^2 = 0$, i.e. $y_1 = y_2$. □

This theorem establishes the method used in solving the problem in Example 87. However, it applies only to L_2 spaces. Thus, it is limited to banal space-time, or geometrically Euclidean spaces, in order to be applicable to various other measurable connections of perceptions. Therefore, we need to, and circumstances allow us to, further elaborate it.

To begin the continuation, let us note that the functional $f = \langle x, y\rangle$ consists of *dual vectors* x and y. This means that each of them is a conjugate transpose (or adjoint for short) type of the other. If quantum operators are self-adjoint (they do not change by adjoint), their vectors are not, so this at least small difference of state exists even in the largest sums. In tensor calculus, this is the difference between contravariant $|y\rangle$ and *covariant coordinates* $\langle x|$ whose product $\langle x, y\rangle$ is an *invariant* (a quantity that does not change the choice of coordinates).

2.4.3 Diversity

A vector space of p-norm, $p \ge 1$, has a dual of q-norm, where $1/p + 1/q = 1$. When using different norms, we treat the "same" space X differently, as L_p and L_q. The dual of L_1 is L_∞ and vice versa, the norms:

$$\|x\|_1 = |\xi_1| + |\xi_2| + ... + |\xi_n|, \quad \|x\|_\infty = \max_{1\le k\le n} |\xi_k|,$$

when dealing with sequences $x = (\xi_1, \xi_2, ..., \xi_n)$ of length n, or if we work with continuous functions $x = x(t)$ on a given interval $t \in (a, b)$:

$$\|x\|_1 = \int_a^b |x(t)|\, dt, \quad \|x\|_\infty = \sup_{a\le t\le b} |x(t)|.$$

The dual of L_2 is L_2, norms:

$$\|x\|_2 = \left(\sum_{k=1}^n |\xi_k|^2\right)^{1/2}, \quad \|x(t)\|_2 = \left(\int_a^b |x(t)|^2\, dt\right)^{1/2}.$$

However, although the dual spaces of L_2 are of equal norms, in the dot product it does not matter which vector is from which dual. For example, if $x = (1, i)$ and $y = (i, 1)$ with imaginary unit $i^2 = -1$, we calculate:

$$\langle x, y\rangle = 2i, \quad \langle y, x\rangle = -2i, \quad \|x\|_2 = \|y\|_2 = \sqrt{2}.$$

The mutual experiences of the states of x and y, as subjects of communication, are not equal. In the physical macro world, such differences caused by complex numbers are relatively smaller, but they do not disappear completely. These differences in vectors (states) are only greater in the case of p-norms when $p \ne 2$, to which are then added different perceptions of the norms (intensity) of phenomena from the point of view of different observers.

In the case of bounded (otherwise denoted by m) infinite sequences ($n \to \infty$) then the space denoted by ℓ_p, the previous form of Riesz's theorem (88) will read similarly. For example, in the real space (4.2. Paragraph) it would be the following paragraph without conjugation.

Example 89. *A bounded linear functional f on the space ℓ_1 is:*

$$\forall x = (\xi_\nu) \in m, \quad f(x) = \sum_{\nu=1}^\infty \xi_\nu \eta_\nu^*, \quad y = (\eta_\nu) \in m.$$

The functional f on ℓ_1 corresponds to a single-valued point y in m, the space of infinite bounded sequences.

We can assume that we are immersed in some infinities that we do not see and, based on the Borel-Kantelli lemma, that we always communicate only with the finitudes of the space ℓ_1. The set of these finitudes for the subjects, y and an arbitrary some x of this example, does not have to be the same. This is an additional increase in the differences in the perceptions of the subjects of the spaces L_1 (finite sequences) and ℓ_1 (infinite sequences), which is why even in spaces of real coefficients (scalar $\Phi = \mathbb{R}$) the participants of communication, or interaction, see the world around them differently. This also occurs with other p-norms.

The following statement with proof (4.6. Stav) is also from Aljancic's book [3], and was posted on the website because that book is not available for some younger colleagues. In the example, I took into account that scalars can be complex numbers ($\Phi = \mathbb{R}$).

Example 90. *A bounded linear functional $f(x)$ on the space $L_p(a,b)$ with parameter $1 < p < \infty$ has the representation:*

$$f(x) = \int_a^b x(t)y^*(t)\,dt,$$

where $y \in L_q(a,b)$ with $1/p + 1/q = 1$ and $\|f\| = \|y\|_q$. The functional $f \in L_p$ uniquely determines the function $y \in L_q$.

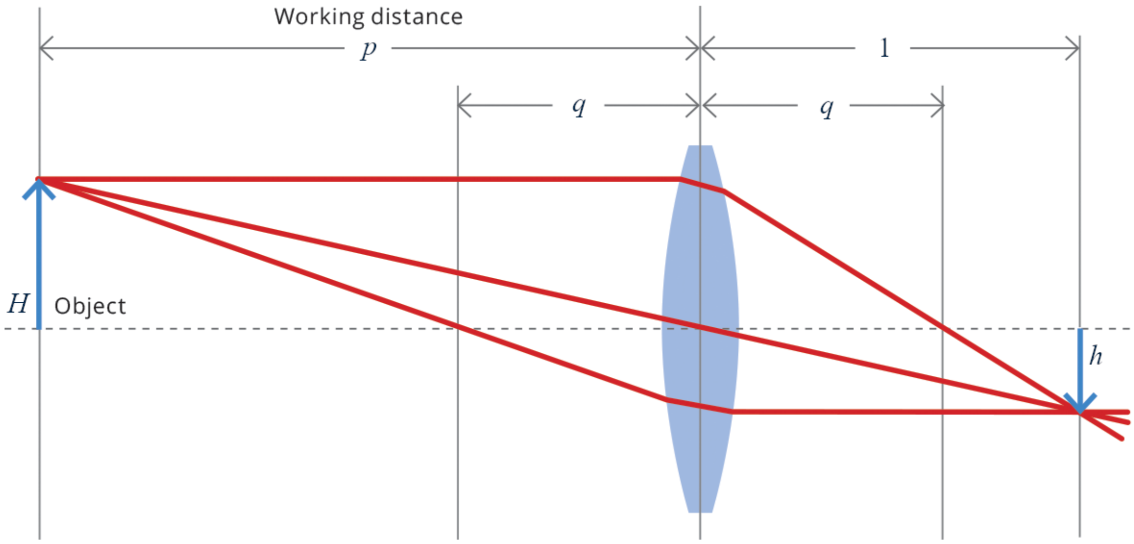
Unlike the previous example, which establishes a similar unique existence of the vector y for infinite finite sequences, this one is about processes (they are also vectors). A process is a unique "subject", which perceives, or communicates, or interacts with the environment from L_p in a unique way, but as a space L_q. What such a subject sees is not what it is, not only because of the differences in the definitions of the two spaces, or because of the different finitudes available to them, but also as a simple consequence of "uniqueness". If some subject saw exactly what "it is", it would be special, absolute.

In the figure, we see a lens that collects light scattered by an object H and creates its image h on a light-sensitive "sensor", usually based on CCD or CMOS (Optics basics). Here, the "distance" p from the lens to the object and the "focal length" q are proportionally calibrated so that the distance from the lens to the image is unity. Thus:

$$\frac{1}{p} + \frac{1}{q} = 1$$

the basic equation of an optical system. When we divide it by the focal length in millimeters, we get its form that photographers work with.

I used this well-known equation from optics to explain conjugation (adjoint, dual space, or contravariance) to a question I asked, roughly: "Are you claiming that the subject achieves its uniqueness by perceiving objects from its own, dual, world, and that therefore what we see is not actually what it is?"You can find part of that otherwise lengthy answer at the link.



The painting resembles the original, but it is not. The "little things"of the internal structure, the weight of the object, the behavior of the substance in contact with solutions or acid, perhaps about the absorption of light and the photoelectric effect if it is a metal plate, are missing. Some animals have poor eyesight, others see differently, only they hear certain sound frequencies, or have a stronger sense of smell. We have fantasies that enrich still life, and we even imagine that we perceive more than there could be. Still life does not even have that.

As in Plato's cave shadow world in the dialogue "The Republic", no subject can know what exactly is happening out there, how another observer, or participant in the interaction, sees it "the same", so we guess, and step by step as we guess, in an absurd way, we are further and further from the experience of ourselves and that physical reality. Mathematics and science, with their inevitable use of untruth as a tool, are not authentic images of the objects of their study.

2.4.4 Convergence

A vector space is defined by four axioms in the section on vector multiplication. In particular, if the first one is satisfied, $\lambda(x+y) = \lambda x + \lambda y$, we know that we have vectors x and y, and the other three are more concerned with scalars, basic properties of numbers such as complex numbers, and this λ. Typical vectors are sequences, say $x = (\xi_1, \xi_2, \dots)$ and $y = (\eta_1, \eta_2, \dots)$, whether they are finite or infinite. It is important that the latter are *convergent* (that the sums of their coefficients are finite), or at least that they are *bounded* sequences of numbers (ξ_ν), which means that $|\xi_\nu| \le M$ for some given number $M > 0$ and for all terms of the sequence. Then:

$$\lambda(x+y) = (\lambda\xi_1 + \lambda\eta_1, \lambda\xi_2 + \lambda\eta_2, \dots) = (\lambda\xi_1, \lambda\xi_2, \dots) + (\lambda\eta_1, \lambda\eta_2, \dots) = \lambda x + \lambda y$$

is the meaning of multiplying sequences by a scalar. There are three common types of *metric spaces* of infinite sequences, and we avoid verbosity by using their short names.

Space $\ell = \ell_1$. Points x of the metric space ℓ are infinite sequences of numbers (ξ_ν) such that the sequence $\sum_{\nu=1}^{\infty} |\xi_\nu|$ converges. In ℓ the distance is introduced by

$$d(x,y) = \sum_{\nu=1}^{\infty} |\xi_\nu - \eta_\nu|.$$

Space m. Points x of the metric space m are bounded sequences of numbers. The distance in m is introduced by

$$d(x,y) = \sup_{1\le\nu<\infty} |\xi_\nu - \eta_\nu|.$$

Space c. Points x of the metric space c are convergent sequences (ξ_ν), and the metric is introduced as in m.

Space c_0. The points x of the metric space c_0 are sequences (ξ_ν) that converge to zero, and the distances are defined as in m.

These sequences are usually accompanied by continuity, and then the space $C[a,b]$ of continuous functions $x(t)$ on the interval $b - a < \infty$, with the distance

$$d(x,y) = \max_{a\le t\le b} |x(t) - y(t)|.$$

The space $C_1[a,b]$ is of continuous functions with the distance

$$d(x,y) = \int_a^b |x(t) - y(t)|\, dt.$$

The space $B[a,b]$ is of bounded functions with the distance

$$d(x,y) = \sup_{a\le t\le b} |x(t) - y(t)|.$$

Bounded functions do not have to be continuous.

When the coefficients of the sequences $x = (\xi_\nu)$ and $y = (\eta_\nu)$ are probabilities, they need not be distributions (of a complete set of independent events) and the question of their convergence becomes important. When we work with information, say logarithms of probabilities, then the question of boundedness is important. Let us consider these by comparing them with the properties of *free particle*, one of the most famous states of quantum mechanics today.

Quantum physics distinguishes a "free particle" from one that is not confined in space, such as a "particle in a box." The basic case is when a particle does not encounter a potential anywhere in space and is along one dimension. It is free to go anywhere in that one dimension without any obstacles:

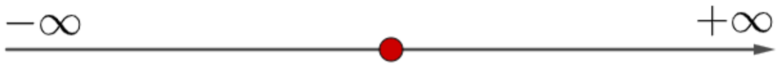
Since the potential is time-independent, we use Schrödinger's time-independent equation to find the *wave function* of a free particle:

$$\hat{H}\psi(x) = E\psi(x).$$

Hamiltonian The operator for this system is:

$$\hat{H} = -\frac{h^2}{8\pi^2 m}\frac{d^2}{dx^2}, \quad -\infty < x < \infty.$$

$$U(x) = 0 \qquad -\infty < x < +\infty$$



The Schrödinger equation for such a problem gives:

$$-\frac{h^2}{8\pi^2 m}\frac{d^2}{dx^2}\psi(x) = E\psi(x),$$

$$\frac{d^2}{dx^2}\psi(x) + \frac{2Em}{(h/2\pi)^2}\psi(x) = 0,$$

$$\frac{d^2}{dx^2}\psi(x) + k^2\psi(x) = 0, \quad k^2 = \frac{2Em}{\hbar^2}.$$

where $\hbar = h/2\pi = 1.054571817... \times 10^{-34}$ Js is Planck's reduced constant. The general solution of this differential equation is:

$$\psi(x) = Ae^{ikx} + Be^{-ikx}, \quad i^2 = -1. \tag{2.22}$$

The wave function must be "bound". It cannot diverge, it cannot take infinite values when x tends to minus or plus infinity. Therefore, the energy must be either positive or zero, $E \ge 0$. For negative energies, $k = \sqrt{2mE}/\hbar$ would be imaginary, so the exponent would be real and diverge at least in one of its two sums. Of the two sums of this wave function:

$$Ae^{ikx} = A\psi_+(x), \quad Be^{-ikx} = A\psi_-(x),$$

the first represents the motion of the particle in the $+x$ direction, and the second the motion in the $-x$ direction.

It turns out that the energy of a free particle is not quantized. It can assume any positive energy. The allowed energies of a free particle are continuous, just as they are in classical mechanics. Since the particle is free, nothing can change its direction of motion. This would require a force, and therefore a potential. Therefore, the particle can only move in one direction along a given path, which means that either A or B is zero.

So, the wave function for a free particle is one of two:

$$\psi_+(x) = Ae^{i\sqrt{2mE}x/\hbar}, \quad \psi_-(x) = Be^{-i\sqrt{2mE}x/\hbar}. \tag{2.23}$$

If we determine the values of these constants, A and B, we use the *Born's law* probability of finding the wave function:

$$\Pr(x)dx = |\psi(x)|^2 dx = \psi^*(x)\psi(x)dx = A^*Ae^{-ikx}e^{ikx} = |A|^2 dx.$$

In this case, the probability of finding the particle between x and $x + dx$ does not depend on x (it is the same everywhere) and is $\Pr(x)dx = |A|^2 dx$, so we say that the particle's position is completely uncertain.

Let's stop here and briefly step outside quantum mechanics. Probability involves fictitious *Hartley information*, given:

$$\text{Inf}(x) = -\ln \Pr(x) = -\ln |\psi(x)|^2 =$$

$$= -\ln \psi^*(x)\psi(x) = -\ln \psi^*(x) - \ln \psi(x)$$

$$= -[\ln A^* - i(kx + 2n_1\pi)] - [\ln A + i(kx + 2n_2\pi)], \quad n_1, n_2 \in \mathbb{Z},$$

$$\text{Inf}(x) = -\ln |A|^2 + 2in\pi,$$

with an arbitrary integer $n = n_1 - n_2 \in \mathbb{Z}$. Complex logarithm is an extension of the real to complex numbers other than zero. The Hartley information of a quantum state is in this sense a "quantum state" itself.

Quantum measurement is a monitored interaction of a closed quantum system with its environment from which the state of the quantum system being measured can be determined. A quantum measurement is given by a set of operators[6], e.g. $\hat{A}_m$, which act on the state space and the probability of measuring m from the state ψ of the system is:

$$\Pr(m) = \langle\psi|A_m^\dagger \hat{A}_m|\psi\rangle, \tag{2.24}$$

and it is a matter of completeness of the distribution that the sum of all measurement results is one:

$$\sum_m \Pr(m) = \sum_m \langle\psi|\hat{A}_m^\dagger \hat{A}_m|\psi\rangle = 1.$$

The state of the quantum system after the measurement is:

$$|\psi'\rangle = \frac{\hat{A}_m|\psi\rangle}{\sqrt{\Pr(m)}}.$$

In short, a measurement is a scalar product $\langle\psi', \psi'\rangle$, where $|\psi\rangle = \hat{A}|\psi\rangle$. Therefore, the ordinary, let's call it "uncontrolled", interaction of this and another quantum system can be interpreted in general as a scalar product $\langle\phi', \psi'\rangle$. This brings us back to the previously mentioned coefficients of the series, in general as $x = (\xi_\nu)$ and $y = (\eta_\nu)$, to the questions of convergence of their product.

Theorem 91. *A necessary and sufficient condition for the series $\sum_{\nu=1}^\infty \xi_\nu \eta_\nu^*$ to (absolutely) converge for every convergent sequence (ξ_ν) is that the series $\sum_{\nu=1}^\infty \eta_\nu$ converges.*

In other words, the sequence $\langle x, y\rangle = \sum_{\nu=1}^\infty \xi_\nu \eta_\nu^*$ converges for every sequence $x = (\xi_\nu) \in c$ if and only if $y = (\eta_\nu) \in \ell$. Note that c and ℓ are dual spaces of infinite sequences.

Proof. The condition is sufficient, because from:

$$\sum_\nu^\infty |\xi_\nu \eta_\nu^*| \le \sup_{1\le\mu<\infty} |\xi_\mu| \cdot \sum_{\nu=1}^\infty |\eta_\nu| = \|x\|_c \cdot \|y\|_\ell < +\infty$$

[6]This is the postulate of quantum measurement.

it follows that the series $\sum_{\nu}^{\infty} \xi_\nu \eta_\nu^*$ converges absolutely.

The condition is necessary. Since the point $y_n = (\eta_1, \eta_2, ..., \eta_n, 0, 0, ...) \in \ell$ for each n, it is:

$$f_n(x) = \sum_{\nu=1}^{n} \xi_\nu \eta_\nu^*, \quad x = (\xi_\nu),\ n = 1, 2, ...,$$

a sequence of bounded linear functionals on the space c and

$$\|f_n\| = \|y_n\|_c.$$

The left side of the equality is the norm of the operator, the right side is the norm of the sequence. By assumption, there exists

$$\lim_{n\to\infty} f_n(x) = \sum_{\nu=1}^{\infty} \xi_\nu \eta_\nu^*$$

for each $x \in c$, so the sequence of norms $\|f_n\|$ is bounded, i.e. $\|f_n\| \le M$. Hence

$$\|y_n\|_\ell = \sum_{\nu=1} |\eta_\nu| \le M$$

for each $n = 1, 2, ...$ which means that the series $\sum_{\nu=1}^{\infty} |\eta_\nu|$ converges, i.e. that $y = (\eta_\nu) \in \ell$. □

Other theorems of this type are proved similarly. A necessary and sufficient condition for the series $\sum_{\nu=1}^{\infty} \xi_\nu \eta_\nu^*$ to converge for every sequence $(\xi_\nu) = x \in \ell_p$ $(p > 1)$ is that the second sequence from the dual space is $(\eta_\nu) = y \in \ell_q$ when $1/p + 1/q = 1$. This property of convergence also holds for extreme duals. A necessary and sufficient condition for the series $\sum_{\nu=1}^{\infty} \xi_\nu \eta_\nu^*$ to converge for every sequence $(\xi_\nu) = x \in \ell$ is that the sequence $(\nu_\eta) = y \in m$. These properties also apply to integrals.

Theorem 92. *A necessary and sufficient condition for the integral $\int_a^b x(t)y^*(t)\ dt$ to exist for every function $x \in L_p$ $(p > 1)$ is that the function $y \in L_q$ where $1/p + 1/q = 1$.*

Proof. That the condition is sufficient follows from Holder's inequalities. To show that the condition is also necessary, let us consider the sequence of functions

$$y_n(t) = \begin{cases} y(t), & |y(t)| \le n, \\ n, & |y(t)| > n. \end{cases}$$

Since for every fixed n, $y_n(t)$ is a bounded function, then $y_n \in L_q$, for every $q > 1$, if the interval is finite and thus for that p when $1/p + 1/q = 1$. Then

$$f_n(x) = \int_a^b x(t) y_n^*(t)\ dt$$

represents a sequence of bounded linear functionals on L_p and

$$\|f_n\| = \|y_n\|_q.$$

Since, by assumption, there exists

$$\lim_{n\to\infty} f_n(x) = \int_a^b x(t)y^*(t)\,dt, \quad \forall x \in L_p$$

then the sequence of norms $\|f_n\|$ is bounded. This means that

$$\|y_n\|_q \le M \quad \forall n = 1, 2, , \dots$$

Letting $n \to \infty$, from Fatou's Lemma, it follows that $\|y\|_q \le M$, i.e. $y \in L_q$. □

This theorem is also valid when $p = 1$, so $x \in L$ and $y \in M$.

2.4.5 Fourier series

From the wave function of a free particle (2.22), we get that the probability of its finding on the x-axis is $\Pr(x) = |A|^2 dx$, that it is the same everywhere, and that the position of the particle is completely uncertain. There is no point in space to which the particle gives preference over any other point in space. This wave function cannot be normalized to a unit value; we say normalized, but this is not necessary for unbound particles.

The function (2.22) thus became an invalid wave function for the postulates of quantum mechanics. However, the settings could be changed to be conditions only for bound states. A free particle is an unbound state, and it became a special case. A *Bound state* arises[7] when the particle is conditioned by the potential $U(x)$. When the particle is localized, it is true: if $x \to \pm\infty$ then $\psi(x) \to 0$. This does not apply to a free particle.

The position of a free particle is completely uncertain. As strange as this situation is, it is in accordance with the uncertainty principle

$$\Delta x \Delta p \ge \frac{h}{4\pi},$$

because the (linear) momentum of the particle is completely determined, certain:

$$\sigma_{p_x}^2 = \langle p_x^2 \rangle - \langle p_x \rangle^2 = 0$$

and hence $\Delta p_x = 0$. So, the momentum of a free particle is completely certain, and the position is completely uncertain (infinite standard deviations).

I remind you, I have this conclusion in the equivalence of physical action ($\Delta E \Delta t$) and information when the law is universal ($\Delta t \to \infty$) and therefore energy-free ($\Delta E \to 0$) so that the action remains constant. From this follows the conclusion that it is the slower flow of time that is attractive, which is at the heart of minimalism. The attraction of the slower flow of time, or movement by inertia, of less action, less uncertainty, order, and more probable outcomes are related phenomena.

[7] Do not confuse the meaning of bound/unbound function and bound/unbound state.

Just as a binary *bit* (binary digit) is the unit of information in traditional computing, a *qubit* (quantum bit) is the fundamental unit of information in quantum computing. A classical binary bit can have only one binary value, such as 0 or 1, which means it can only be in one of two possible states. However, a qubit is a superposition from quantum mechanics and achieves a linear combination of these two states. It can represent 0 or 1, but also any proportion of 0 and 1 in the superposition of both states, with a certain probability of being 0 and a certain probability of being 1. Here is an example of a projective measurement with such.

Projective measurement concerns deterministic measurement processes of quantum systems when a set of orthogonal operators, known as projectors, are used to observe the outcome. These measurements are repeatable and the outcome is deterministic.

Example 93. *Projective measurement of a qubit in the ψ state on a computational basis.*

Description.

$$|\psi\rangle = \begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \alpha\begin{pmatrix}1\\ 0\end{pmatrix} + \beta\begin{pmatrix}0\\ 1\end{pmatrix} = \alpha|0\rangle + \beta|1\rangle \qquad |\alpha|^2 + |\beta|^2 = 1.$$

The measurement operators are:

$$\hat{A}_0 = |0\rangle\langle 0| = \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}, \quad \hat{A}_1 = |1\rangle\langle 1| = \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}.$$

Measurement probabilities:

$$\Pr(0) = \langle\psi|A_0^\dagger A_0|\psi\rangle = \alpha^*\alpha\langle 0,0\rangle = |\alpha|^2,$$

$$\Pr(1) = \langle\psi|A_1^\dagger A_1|\psi\rangle = \beta^*\beta\langle 1,1\rangle = |\beta|^2.$$

State after measurement:

$$\frac{A_0|\psi\rangle}{\sqrt{\Pr(0)}} = \frac{\alpha|0\rangle}{\sqrt{|\alpha|^2}} = \frac{\alpha}{|\alpha|}|0\rangle, \quad \frac{A_1|\psi\rangle}{\sqrt{\Pr(1)}} = \frac{\beta|1\rangle}{\sqrt{|\beta|^2}} = \frac{\beta}{|\beta|}|1\rangle.$$

Re-measuring a state, after the first measurement, yields the same state as the initial measurement but then with unit probability. □

We note that the measurement resulted in the discharge of uncertainty. This is typically stated in computer terms and means that the electron only gains its previous trajectory after observation (the founders of quantum mechanics noticed this with amazement), and now we say that because the measurement transfers its uncertainty to the apparatus, after which its trajectory remains with greater certainty. Second, the vectors $|0\rangle$ and $|1\rangle$ are orthonormal.

The scalar product of two functions f_1 and f_2 on an interval (a,b) is the number

$$\langle f_1, f_2\rangle = \int_a^b f_1(t) f_2^*(t)\, dt.$$

The functions f_1 and f_2 are *orthogonal* on a given interval when $\langle f_1, f_2\rangle = 0$. A set of functions is orthogonal on a given interval if every pair of those functions is orthogonal on the given interval, and it is orthonormal if the functions are also normalized.

Example 94. *The set of cosines*

$$\frac{1}{\pi}, \frac{1}{\pi}\cos t, \frac{1}{\pi}\cos 2t, \frac{1}{\pi}\cos 3t, ...$$

on the interval $(-\pi, \pi)$ *is orthonormal.*

Proof. Namely:

$$\int_{-\pi}^{\pi} \cos mt \cos nt\ dt = \int_{-\pi}^{\pi} \frac{1}{2}[\cos(m-n)t + \cos(m+n)t]\ dt =$$

$$= \frac{1}{2}[\frac{\sin(m-n)t}{m-n} + \frac{\sin(m+n)t}{m+n}]\Big|_{-\pi}^{\pi} = 0,$$

because $m-n$ and $m+n$ are integers and the sines of such products with π are zero. But when $m = n$, we calculate:

$$\int_{-\pi}^{\pi} \cos^2 nt\ dt = \int_{-\pi}^{\pi} \frac{1}{2}(1 + \cos 2nt)\ dt = \frac{1}{2}t\Big|_{-\pi}^{\pi} + \frac{1}{2}\frac{sin2nt}{2n}\Big|_{\pi}^{\pi} = \pi$$

so the set is orthonormal because the elements are divided by π. □

By changing $t \to \pi/2 - t$ the cosines become sines, so the resulting sequence of sines is equally orthonormal (orthogonal and normalized). When $x(t)$ is an integrable function of period 2π, which means that $x(t + 2\pi) = x(t)$ for each t, then we can define the *Fourier coefficients* of x and the Fourier series $f(t)$ as:

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} x(t)\cos nt\ dt, \quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} x(t)\sin nt\ dt, \quad n = 1, 2, ...$$

$$f(t) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n \cos nt + b_n \sin nt). \tag{2.25}$$

The sequence of Fourier coefficients (a_n) is a sequence of bounded linear functionals on $L(-\pi, \pi)$. Namely, since:

$$|a_n(x)| \le \frac{1}{\pi}\int_{-\pi}^{\pi} |x(t)|\ dt = \frac{1}{\pi}\|x\|_L$$

then the sequence of their norms $\|a_n\|$ is bounded ($\le 1/\pi$).

Example 95. *Let us show that* a_k *and* b_k *in the Fourier series*

$$f(t) = \frac{1}{2}a_0 + \sum_{k=1}^{\infty}(a_k \cos nt + b_k \sin nt)$$

are exactly the Fourier coefficients.

Solution. Multiply the given equation by $\cos(nt)$, for $n \geq 1$ and integrate from $-\pi$ to π, assuming that sum-by-sum integration is allowed:

$$\int_{-\pi}^{\pi} f(t)\cos nt\, dt = \frac{1}{2}a_0 \int_{-\pi}^{\pi} \cos nt\, dt + \int_{-\pi}^{\pi} \sum_{k=1}^{\infty} (a_k \cos kt \cos nt + b_k \sin kt \cos nt)\, dt = a_n\pi,$$

$$= \frac{1}{2}a_0 \frac{\sin nt}{n}\Big|_{\pi}^{\pi} + \sum_{k=1}^{\infty}\left(\int_{-\pi}^{\pi} a_k \cos kt \cos nt\, dt\right) + \sum_{k=1}^{\infty}\left(\int_{-\pi}^{\pi} b_k \sin kt \cos nt\, dt\right)$$

$$= 0 + \pi a_n + \sum_{k=1}^{\infty} \int_{-\pi}^{\pi} \frac{b_k}{2}[\sin(k-n)t + \sin(k+n)t]\, dt = \pi a_n$$

because all the other sums are zero. Similarly, multiplying the given equation by $\sin(nt)$ and integrating, we find

$$\int_{-\pi}^{\pi} f(t)\sin nt\, dt = \pi b_n.$$

Hence a_k and b_k are exactly the Fourier coefficients (2.25), with $x(t) = f(t)$. $\square$

Many functions $f(t)$ can be represented using Fourier series (Convergence), which means that the Fourier coefficients are zero sequences, i.e. if $n \to \infty$ then both $a_n, b_n \to 0$. Another way to represent them is based on Euler's formula $e^{int} = \cos nt + i \sin nt$. Hence:

$$\cos nt = \frac{1}{2}(e^{int} + e^{-int}), \quad \sin nt = \frac{1}{2}(e^{int} - e^{-int}),$$

where we have an orthonormal system $\{v_n = \frac{1}{\sqrt{2\pi}}e^{int}\}$. Namely:

$$\langle v_m, v_n\rangle = \int_{-\pi}^{\pi} \frac{e^{imt}}{\sqrt{2\pi}}\frac{e^{-int}}{\sqrt{2\pi}}\, dt = \begin{cases} 1, & m = n \\ 0, & m \neq n \end{cases}$$

The Fourier series expansion of the function from the example then becomes:

$$f(t) = \sum_{n=-\infty}^{+\infty} c_n e^{int}, \quad c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-int}\, dt. \tag{2.26}$$

More generally, when $v_1(t), v_2(t), \ldots$ are mutually orthogonal eigenfunctions, and we expand the function $f(t)$ on the interval (a, b) using these eigenfunctions, we have:

$$f(t) = \sum_{n=1}^{\infty} A_n v_n(t), \tag{2.27}$$

with coefficients:

$$A_n = \frac{\langle f, v_n\rangle}{\langle v_n, v_n\rangle} = \frac{\int_a^b f(t)v_n^*(t)\, dt}{\int_a^b |v_n(t)|^2 dt}.$$

A special case of this general expression is the previous, complex form of the Fourier expansion, which we can check by setting $v_n(t) = e^{int}$.

Example 96. *Let's find the Fourier expansion of the function $f(t) = t^2$ in the interval* $(0, 2\pi)$.

Solution. We use trigonometric formulas (2.25). Instead of the interval $(-\pi, \pi)$ we use $(0, 2\pi)$ and calculate:

$$a_0 = \frac{1}{\pi}\int_0^{2\pi} t^2\,dt = \frac{8\pi^2}{3},$$

$$a_n = \frac{1}{\pi}\int_0^{2\pi} t^2 \cos nt\,dt = \frac{4}{n^2}, \qquad n \neq 0,$$

$$b_n = \frac{1}{2\pi}\int_0^{2\pi} xt^2 \sin nt\,dt = -\frac{4\pi}{n},$$

$$f(t) = \frac{4\pi^2}{3} + \sum_{n=1}^{\infty}\left(\frac{4}{n^2}\cos nt - \frac{4\pi}{n}\sin nt\right).$$

This is valid, $f(t) = t^2$, for $0 < t < 2\pi$. At the boundary points of the interval, for $t = 0$ and $t = 2\pi$, the series converges to $2\pi^2$. □

The advantage of expanding in Fourier series is, for example, if you need to prove the (Basel problem) that

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \dots = \frac{\pi^2}{6}.$$

Namely, at the point $t = 0$ the Fourier series converges to that limit, so we have

$$\frac{4\pi^2}{3} + \sum_{n=1}^{\infty}\frac{4}{n^2} \to 2\pi^2, \qquad t \to 0,$$

and hence the desired result.

The Fourier series is commonly used to describe a periodic signal in terms of cosine and sine waves. It helps engineers translate any periodic signal into a combination of sine and cosine. However, we have a different ambition here.

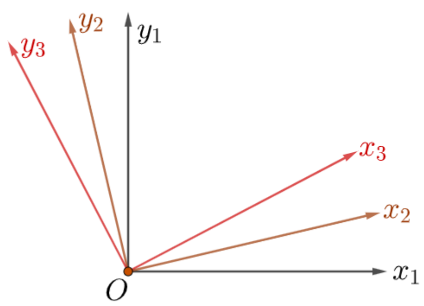
We have seen the transition from a simple trigonometric expression (2.25) for the Fourier series expansion of a function to a complex one (2.26), and then to a general one (2.27). This third one is targeted and actually represents the function as a vector using an orthogonal basis of that vector space. For example, the possibility of decomposing a given function into various graph shapes, not just fragments of a sinusoid, means that the smallest parts of the function have each of these shapes, or rather, they do not have any particular shape.

These interpretation functions are of vectors or mappings and then states or processes, thereby revealing to us strange properties of reality. The choices of various orthogonal bases of space are, in principle, endless, arbitrary, and not in any essential way special, but with equally extreme equality, they describe the same given function in different ways. They remind us of the changes of the same amount of energy from form to form when we speak of the perceptions of the same object by different subjects.

The third thing that can inspire us while translating regular, broken, and other lines into Fourier series is the irregularity of their details. Series that can start with any of the shapes, i.e., the absence of a special shape, turn into regularities, or vice versa. We will say that nature in the physical world does not build straight paths or pure triangular or rectangular shapes, but they are equal to the ones familiar to us, like the jungle. The thing is only that the latter are so numerous that the former are negligible and invisible.

2.4.6 Counting

Let us pause briefly to enumerate the orthonormal bases of a given infinite-dimensional vector space. We want to see the uncountably infinite possibilities of representing a function in the form of infinite series, such as generalized Fourier series or similar ones, and then to understand some of the consequences of such knowledge.



The figure on the left shows various rectangular systems of $Ox_\nu y_\nu$ abscissas and ordinates, always in the same plane, perpendicular to the applicate, the z-axis. With this in mind, let's imagine an infinite series of coordinate axes, each of them perpendicular to all the others. As in the figure, let's change adjacent axes two by two at will, so that they remain mutually perpendicular and in a plane that is perpendicular to all the others. In this way, we achieve different orientations of each of the infinite set of coordinate axes, while their overall system is always orthogonal.

Let us now return to the expansion (2.27) of an arbitrary function $f(t)$ in the "Fourier series" with respect to mutually orthogonal vectors $v_n(t)$. The list of sums of such an expansion in series contains a countably infinite sequence $A_1, A_2, A_3, ...$ of coefficients, mutually unique (by bijection) associated with the terms of the sum. By changing the basis vectors v_n, these coefficients change in very different ways. But if an infinite subset of them could take only two different values, this countably infinite sequence would become an uncountably infinite set of possibilities.

The result is as in the case of decimals of real numbers from the interval (0, 1) which we write in countably infinite sequences $0, d_1 d_2 d_3 ...$ of cardinal number $\aleph_0$, uncountably infinite real numbers of cardinal number $\mathfrak{c}$. The supposed decimals can also be just two digits 0 or 1 of the binary representation of the real number, with the same effect.

Uncountable infinity, simply put, means that an infinite sequence is not enough for all possibilities. When "real"is any of the following: 1. to be known by observation; 2. to be sustained in some way; 3. if it is true – this invokes the diversity of the very concept of "reality"as well as its relativity. There are no two subjects (living or non-living) who see their surroundings in the same way. With all this, this third

one is of particular interest to us now. Let's say, where are those possibilities that could not be sequenced?

When the object A_1 can be observed by A_2, this one by A_3 and so on, that one by A_{n-1}, then they are all mutually "real". This is more precisely stated in (1). However, some sustainability of that chain is implied, so we also have (2), and then there is consistency (3). When we have two equal possibilities, for a "tail" or "head" to fall, and if one of them happens to us, then the other has gone into a "parallel reality". For the successors of that other, all events given to them will be just as observable (1), sustainable (2) and true (3) as ours. We will have a common past, but different futures.

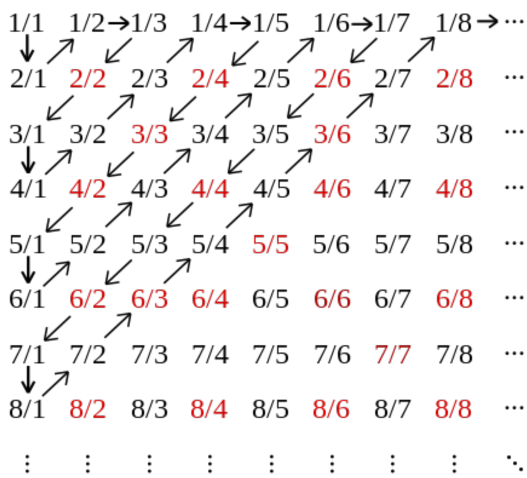
This is the answer to the question: "Where can we put uncountably infinite possibilities, when we are dealing with the most countable infinity of events?" And at the same time preserve the objectivity of chance. The aforementioned most interesting third option has several other interesting pitfalls. Before that, let's review once again what we count and how we count.

The set of natural numbers ($\mathbb{N}$) is "countably infinite" and every set with which it can establish a *bijection* (a two-sided one-to-one association) is also of the same *cardinality* $\aleph_0$, is countably infinite. Therefore the "number" of integers ($\mathbb{Z}$) is also $\aleph_0$.

$$\begin{array}{lcccccccc} \mathbb{Z}: & 0 & 1 & -1 & 2 & -2 & 3 & -3 & \dots \\ \mathbb{N}: & 1 & 2 & 3 & 4 & 5 & 6 & 7 & \dots \end{array} \tag{2.28}$$

It is possible to establish a bijection of rational numbers ($\mathbb{Q}$) with natural numbers. We usually do this in two steps, in the manner of Cantor (Georg Cantor, 1845-1918), who is the founder of set theory and this idea of measuring the immeasurable.

The matrix on the right lists positive fractions with increasing numerators in rows and increasing denominators in columns. We count them diagonally, following the arrows and skipping all (red) values that are repeated. This gives us a bijection of natural numbers and all positive fractions. Then, we insert their negative values between the erms of such a sequence. The result is, as in counting integers (2.28), $\text{kard}(\mathbb{Q}) = \text{kard}(\mathbb{N}) = \aleph_0$.



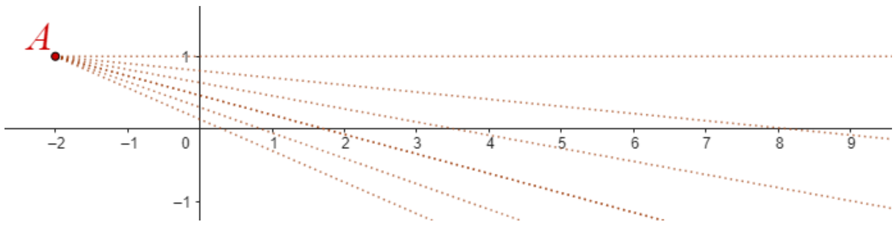
It is said that Cantor struggled for a long time with other sets, until he found the following proof that the real numbers are not countable. They are of higher cardinality than the natural numbers, but as a cautious mathematician he did not decide to call it "aleph-one", but used the name *continuum* and the notation $\mathfrak{c}$.

0	•	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	...
0	•	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	...
0	•	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	...
0	•	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	...
0	•	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	...
0	•	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	...
0	•	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	...
0	•	h_1	h_2	h_3	h_4	h_5	h_6	h_7	h_8	h_9	...

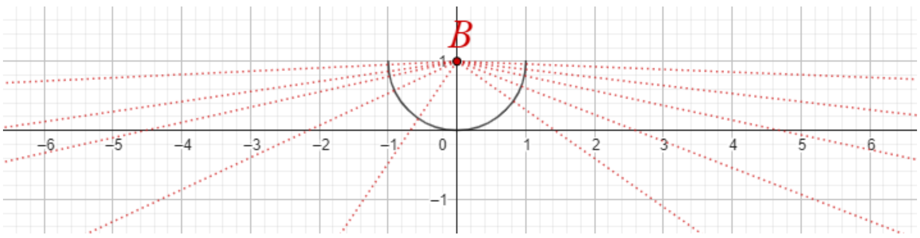
The first step is to prove that the real numbers in the interval (0, 1) are uncountable. Suppose, on the contrary, that the set of real numbers in the interval (0, 1) is countable, and arrange them in an infinite sequence as in the figure on the left. They are all written with zero, comma, and decimal places.

From the first, we replace the first decimal with another digit $a_1' \ne a_1$ and we get a different real number $0, a_1' a_2 a_3 ...$ which is also in the interval (0, 1). From the second, we replace the second decimal ($b_2' \ne b_2$), from the third the third and so on in order, we get a series of new numbers. Then, with the new decimals, we form the real number $x = 0, a_1' b_2' c_3' ... \in (0, 1)$.

The number x differs from the first number of the given sequence by the first decimal place. But it also differs from the second number in the sequence by the second decimal place, from the third by the third decimal place, and in fact from each number in the sequence in the image, the number x differs from each number in its own decimal place. So, this new number is real, it belongs to the same interval, but it is not in the list. This is a contradiction with the statement that all real numbers are listed. Hence the conclusion that the set of positive real numbers less than one is uncountable.



The interval numbers (0, 1), on the image of the vertical axis (ordinate), map the half-lines from point A by bijection to positive real numbers of the abscissa, the horizontal axis, so $\mathrm{kard}(0,1) = \mathrm{kard}\,\mathbb{R}^+ = \mathfrak{c}$. A further bijection is also possible, from point B from position 1 of the ordinate, of a semicircle of arc length π by half-lines to all real numbers of the abscissa. We see it in the following image, and hence $\mathrm{kard}\,\mathbb{R} = \mathfrak{c}$. Zermelo and other followers of Cantor's ideas found infinities less than $\mathfrak{c}$ and greater than $\aleph_0$, also greater than the continuum.



According to this theory of information, we do not doubt that there is a continuum as a kind of "reality"and it is only a matter of discovery where it is and how to

understand it as such. With the accepted theory of sets and its cardinal numbers, we reject the possibility of contradiction in such an expansion of the understanding of reality and we are left with only our (un)openness to novelty as an opportunity in cognition. And the properties of the continuum are only one part of the universal truth.

Concrete phenomena are so utterly true that any evidence of their falsity excludes their existence. Truths are the "tissue" of concreteness without which there would be no point in studying it. We actually accept the abstract laws of the natural world and the natural world as two sides of the same coin, although we are rarely aware of it. The idea of information will unite them into the same type of phenomena.

Physical information is the equivalent of an action ($\Delta E \Delta t$) of the order of magnitude of *Planck's constant* ($h = 6.626\,070\,15 \times 10^{-34}$ J$\cdot$Hz^{-1}) for which the conservation law holds. However, changes in energy (ΔE) over time (Δt) are not constant, and this is where the *principle of minimalism* comes into play. It is reflected in the more frequent outcomes of more probable random events, in the tendency of systems towards less information, generally less communication, a smaller volume of interactions, inertia, the principle of least action in physics. Consistent with all this, the principle of minimalism spontaneously leads systems towards smaller changes in energy. Then $\Delta E \to 0$ means that $\Delta t \to \infty$ so that the product $\Delta E \Delta t$ remains unchanged.

Longer periods of time for the same types of events are relative slowdowns of the flow of time, and they attract. Just as not everything communicates with everything, so we have material differences and different laws to which they gravitate. The avoidance of uncertainty is the emergence of minimalism, the tendency to reduce options, and the surrender of freedom for the sake of security or efficiency, the pursuit of more order, a better order, the path to subjugation. This information theory thus connects laziness with unfreedom and natural laws. Natural phenomena follow the laws of nature, avoiding uncertainty. The laws are irresistible (to the chosen ones) because they are timeless (of unlimited duration).

Unlike the concrete, abstract truths extend further, we say they are omnipresent. Calculating a 3-dimensional functional (primer 87) based on its mapping of three vectors into three numbers, a unique solution is mentioned because the vectors are independent, they span the (3-dimensional) space of the functional. But if there are too few vectors, when they do not encompass the entire space of the functional, then ambiguity, infinity of the solution appears. In this way, I look at the uniqueness of the concrete (subjects of perception) versus the unlimited occurrence of schemes of parts of abstract truths. In short, this is how I understand the differences between Rees's and Skolem's theorems, for now.

This omnipresence ($\Delta t \to \infty$) and omnipresence ($\Delta x \to \infty$) supports this thesis, along with the absence of energy and momentum, due to $\Delta E \Delta t = \Delta p \Delta x$ of the order of Planck's constant, in the operation of abstract laws. This is not its only harmony, but the story is complex and the chances of some inconsistency emerging are still high.

2.4.7 Bessel functions

An unusual addition to the variety are the representations of a given function in the form of infinite series, Fourier or similar, otherwise orthogonal Bessel functions. First of all, they describe the modes of vibration of circular membranes, such as the surfaces of drums. They are used for the analysis of sound waves and vibrations of musical instruments, but also in engineering structures. In cylindrical coordinates, Bessel functions are used to solve problems involving electromagnetic fields, with wave propagation in cylindrical waveguides and in antenna theory. In quantum mechanics, Bessel functions appear in solutions to the Schrödinger equation of a system with cylindrical symmetry.

The parametric Bessel equation and its self-adjoint form:

$$x^2y'' + xy' + (\alpha^2x^2 - n^2)y = 0 \;\Rightarrow\; \frac{d}{dx}(xy') + \left(\alpha^2 x - \frac{n^2}{x}\right)y = 0 \tag{2.29}$$

have two solution groups $J_n(\alpha x)$ and $Y_n(\alpha x)$ but only the first solution has a boundary at the point $x = 0$. The set $J_n(\alpha_k x)$ is orthogonal to the weights $\rho(x) = x$ on the interval $[0, b]$:

$$\int_0^b xJ_n(\alpha_i x)J_n(\alpha_j x)\,dx = 0, \quad \alpha_i \neq \alpha_j. \tag{2.30}$$

The parameters α_k are given by the boundary conditions at $x = b$ with

$$A_2J_n(\alpha b) + B_2\alpha J_n'(\alpha b) = 0.$$

The Fourier-Bessel expansion of the function f in order on the interval $[0, b]$ using Bessel functions is:

$$f(x) = \sum_{k=1}^{\infty} c_k J_n(\alpha_k x), \tag{2.31}$$

where:

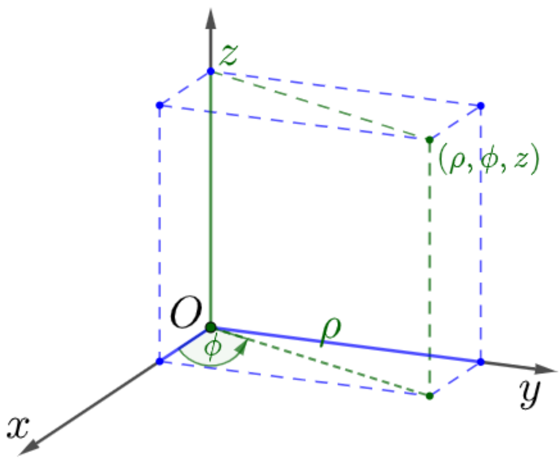
$$c_k = \frac{\int_0^b xJ_n(\alpha_k x)f(x)\,dx}{\|J_n(\alpha_k x)\|^2},$$

and the square of the norm of the function $J_n(\alpha_k x)$ is defined by:

$$\|J_n(\alpha_k x)\|^2 = \int_0^b xJ_n^2(\alpha_k x)\,dx.$$

The figure shows cylindrical coordinates $x = \rho\cos\phi$, $y = \rho\sin\phi$, $z = z$. *Cylindrical symmetry* means independence from cylindrical coordinates ϕ and z when moving on the surface of a cylinder.

Bessel functions are common in problems with cylindrical symmetry. When we see cylindrical symmetry, we usually see "Bessel functions." With spherical symmetry (movements along the sphere), we have "Legendre polynomials," and with Cartesian symmetry (movements along the abscissa, ordinate, and applicate), we have the familiar "sine and cosine."



Example 97. *When* $\alpha = 1$*, Bessel's equation (2.29) is a differential equation of the form*

$$x^2y'' + xy' + (x^2 - n^2)y = 0$$

whose solution is

$$J_n(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\Gamma(n+k+1)} \left(\frac{x}{2}\right)^{2k+n}$$

where

$$\Gamma(z) = \int_0^\infty e^{-t}t^{z-1}\,dt$$

is the gamma function, with a complex number $z \in \mathbb{C}$ *of real part* $\Re(z) > 0$*.*

Solution. The first way is usually (Frobenius Method) to find a solution of infinite order form, which is a possible form for almost any function. The second way is to plug the solution into the equation once the solution is known. Both are computationally long and demanding. □

Gamma function is a generalization of the factorial to complex numbers. For all $n = 0, 1, 2, ...$ it is $\Gamma(n+1) = n!$ because:

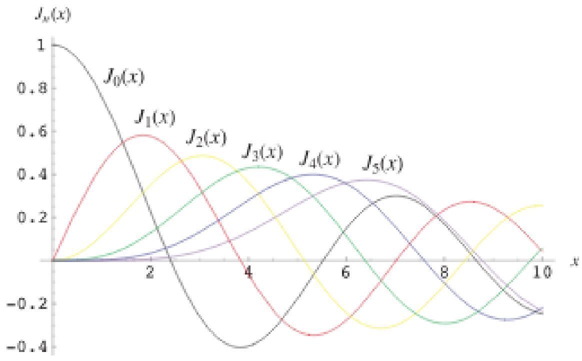
$$\Gamma(n+1) = \int_0^\infty e^{-t}t^n\,dt = -\int_0^\infty t^n\,de^{-t} = -e^{-t}t^n\Big|_0^\infty + \int_0^\infty e^{-t}\cdot nt^{n-1}\,dt = n\Gamma(n).$$

In addition to the above integral, the often mentioned

$$\Gamma(z) = \lim_{n\to\infty} \frac{n!n^z}{z(z+1)(z+2)...(z+n)},$$

Euler's definition of the gamma function.

In the image on the right, we see the first few graphs of the solutions to the Bessel equation given in Example 97, the so-called Bessel functions of the first kind. They are non-singular (with a unique solution) at the origin. They are sometimes called cylinder functions or cylindrical harmonics. The program for drawing these graphs is besel().

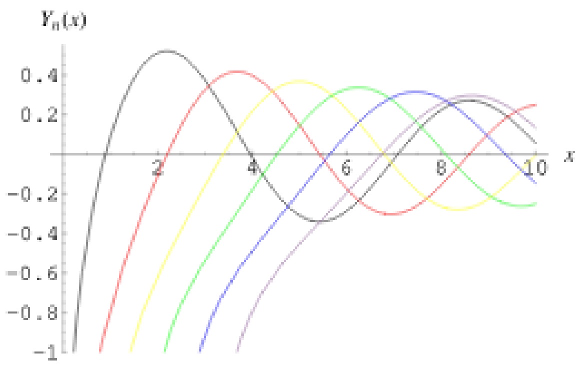


The second solution of the Bessel equation can be found using order reduction once we know the first solution. n does not have to be an integer, and let $\nu \in \mathbb{C}$ be some complex constant. Then

$$Y_\nu(z) = \lim_{n\to\nu} \frac{(\cos n\pi)J_n(z) - J_{-n}(z)}{\sin n\pi}$$

is a Bessel function of the second kind of order ν and argument z. When ν is not an integer, the limit is found by substitution

$$Y_\nu(z) = \frac{(\cos n\pi)J_\nu(z) - J_{-\nu}(z)}{\sin \nu\pi}.$$



Since $Y_\nu(z)$ is a linear combination of solutions of the first kind, this is also a solution. Otherwise, for all integers n, the Bessel function of the second kind Y_n is a solution of the Bessel equation linearly independent of $J_n(z)$. The figure on the left shows the graphs of Bessel functions of the second kind. At the origin, it is singular (a solution that cannot be derived from the general solution of the differential equation). Bessel functions of the second kind are also called Neumann or Weber functions.

The general solution of the Bessel equation from Example 97 is:

$$w(z) = AJ_\nu(z) + BY_\nu(z), \quad A, B \in \mathbb{C}. \tag{2.32}$$

These are general matters that actually interest us less than the question of orthogonality (2.30) of Bessel functions of the first kind.

Example 98. *Let us show that for the solutions of the Bessel equation, a function of the first kind, the equality*

$$\frac{d}{dx}[x^p J_p(x)] = x^p J_{p-1}(x).$$

The parameter $p = n$ does not have to be an integer.

Proof. Starting from the solution of Example 97, we calculate:

$$\frac{d}{dx}[x^p J_p(x)] = \frac{d}{dx}\sum_{k=0}^{\infty}\frac{(-1)^k}{\Gamma(k+1)\Gamma(k+p+1)}\left(\frac{x}{2}\right)^{2n+2p} =$$

$$= \sum_{k=0}^{\infty}\frac{(-1)^k}{\Gamma(k+1)\Gamma(k+p+1)}\frac{2k+2p}{2}\left(\frac{x}{2}\right)^{2k+2p-1}$$

$$= \sum_{k=0}^{\infty}\frac{(-1)^k}{\Gamma(k+1)\Gamma(k+p)}\cdot x^p\left(\frac{x}{2}\right)^{2k+2p-1},$$

and this is $x^p J_{p-1}(x)$. □

Similarly, we find $\frac{d}{dx}[x^{-p}J_p(x)] = -x^{-p}J_{p+1}(x)$, and adding and subtracting the two:

$$J_{p-1}(x) + J_{p+1}(x) = \frac{2p}{x}J_p(x), \quad J_{p-1}(x) - J_{p+1}(x) = 2J_p'(x).$$

Subtracting the second from the first, we also find the equation

$$J_p'(x) = \frac{p}{x}J_p(x) - J_{p+1}(x). \tag{2.33}$$

Example 99. *If αx and βx are two solutions of the Bessel equation, $J_p(\alpha x) = J_p(\beta x) = 0$, let us show that*

$$(\alpha^2 - \beta^2)\int_0^b xJ_p(\alpha x)J_p(\beta x)\,dx = \beta b J_p(\alpha b)J_n'(\beta b) - \alpha b J_p(\beta b)J_n'(\alpha b).$$

Proof. From equation (2.29) and the given conditions, it follows:

$$\frac{d}{dx}\left[x\frac{d}{dx}J_p(\alpha x)\right] + \left(\alpha^2 x - \frac{p}{x}\right)J_p(\alpha x) = 0,$$

$$\frac{d}{dx}\left[x\frac{d}{dx}J_p(\beta x)\right] + \left(\beta^2 x - \frac{p}{x}\right)J_p(\beta x) = 0.$$

We multiply the first by $J_p(\beta x)$, and the second by $J_p(\alpha x)$ and subtract:

$$(\alpha^2-\beta^2)xJ_p(\alpha x)J_p(\beta x) = -J_p(\beta x)\frac{d}{dx}\left[x\frac{d}{dx}J_p(\alpha x)\right] + J_p(\alpha x)\frac{d}{dx}\left[x\frac{d}{dx}J_p(\beta x)\right].$$

We (partially) integrate this in the limits $x \in [0, b]$:

$$(\alpha^2-\beta^2)\int_0^b xJ_p(\alpha x)J_p(\beta x)\,dx =$$

$$= -\int_0^b \frac{d}{dx}\left[x\frac{d}{dx}J_p(\alpha x)\right]J_p(\beta x)\,dx + \int_0^b \frac{d}{dx}\left[x\frac{d}{dx}J_p(\beta x)\right]J_p(\alpha x)\,dx$$

$$= x\frac{dJ_p(\beta x)}{dx}J_p(\alpha x) - x\frac{dJ_p(\alpha x)}{dx}J_p(\beta x)\Big|_0^b$$

$$= \beta b\frac{J_p(\beta x)}{dx}\Big|_{x=b}J_p(\alpha b) - \alpha b\frac{dJ_p(b)}{dx}\Big|_{x=b}J_p(\beta b),$$

because at the lower limit these functions are zero, hence the theorem. □

In many textbooks, the proof of the orthogonality of Bessel functions ends here, because the right-hand side is zero, due to $J_p(\alpha b) = J_p(\beta b) = 0$, so the left-hand side when $\alpha \neq \beta$ the integral must be zero, and this is then (2.30). We simplify the statement in the example (visually) by setting $b = 1$, which does not reduce the generality of the proof.

Example 100. *Let us show that:*

$$\int_0^1 xJ_p(\alpha x)J_p(\beta x)\,dx = \begin{cases} 0, & \alpha \neq \beta \\ \frac{1}{2}[J_{p+1}(\alpha)]^2, & \alpha = \beta \end{cases}$$

where α *and* β *are the roots of* $J_p(x) = 0$.

Proof. Let's start from Example 99 and put $b = 1$. When $\alpha \neq \beta$ and divide the equation by $\alpha^2 - \beta^2$, we have:

$$\int_0^1 xJ_p(\alpha x)J_p(\beta x)\,dx = \frac{-1}{\alpha^2 - \beta^2}[\alpha J_p(\beta)J_p'(\alpha) - \beta J_p(\alpha)J_p'(\beta)].$$

Since α and β are the roots of the equation $J_p(x) = 0$, it follows that $J_p(\alpha) = 0$ and $J_p(\beta) = 0$, so the expression in the square brackets is zero and the above result of the theorem (for $\alpha \neq \beta$) is proven.

When $\alpha = \beta$, let us observe the approximation of the first to the second, $\alpha \to \beta$, which is a fixed root of the equation $J_p(x) = 0$. The right-hand side of the same equality becomes:

$$\lim_{\alpha\to\beta} \frac{0 - \beta J_p'(\alpha)J_p'(\beta)}{-2\alpha} = \frac{1}{2}[J_p'(\alpha)]^2 = \frac{1}{2}[J_{n+1}(\alpha)]^2,$$

this by the definition of the derivation, then by (2.33). Thus, the second part of the theorem (for $\alpha = \beta$) is also proved. □

This proves that the roots of the equation $J_p(x) = 0$, with density or weight $\rho(x) = x$, form an orthogonal set of vectors. This is not new, about the breadth of the representation of functions using various linear combinations of very different operators as a basis of a vector space, but it is interesting. The unusualness of the Bessel functions $J_p(x) = 0$ and especially the use of the density $\rho(x)$ in such a scalar product demonstrate all this unboundedness. The lesson for information theory is that any idea of some absolute reality beyond perception, of something that exists but we try to perceive in vain – would be redundant.

A supposed absolute world that would be beyond the perception of all subjects would not be real. Then, the subject that could perceive such a world would have to last as long as the world itself, because it is unique and unrepeatable in the next moment. So, if there were a subject that perceives such an "absolute" it would itself be timeless, and this again brings us back to the uniqueness of subjects, that is, to the unrepeatability of their perceptions of the world. Then to the unreality of what none of them can perceive anymore.

Finally, natural laws are repeatedly repeatable abstract truths and are not subjects, so that the idea of the world resting on uncertainty (information) is not contradictory. Their indestructible accessibility to subjects also tells us about the (potential) presence of untruth in the composition of their unrepeatable. From Gödel's theorem on the incompleteness of *consistency* (correctness, ultimate truthfulness) of a system, the incompleteness of such a system follows, and further, the differences in the perception of such a system by subjects. In other words, there are no truths that we see equally. But unlike concrete reality, abstract truths can exist as universal truths, and I hope that we now understand why.

On the other hand, from the presence of orthogonal Bessel functions and their like, it is known that there is no shortage of options for multiple perceptions of the world. But also, that without uniqueness there is no multiplicity and vice versa, just as we would not know much about the truth if we did not know how to use lies; moreover, that reality exists thanks to fiction and fiction with the help of reality.

2.4.8 Weight functions

Even on a finite-dimensional one-dimensional vector space, there are many different ways to do (scalar) *inner multiplication*. Choosing one of the scalar products on a vector space X will define the lengths of the vectors and the (cosines) of the angles between the two vectors, and there are infinitely many possible ways, each of which determines a different orthonormal basis for X.

We have seen that the vector $x = (\xi_1, \xi_2, ..., \xi_n)$ can be an n-tuple of numbers from $\mathbb{C}$ that can be mapped into such numbers by linear functionals, where each $f : X \to \mathbb{C}$ single functional can be represented by a unique vector $y = (\eta_1, \eta_2, ..., \eta_k) \in X$ such that:

$$f(x) = \langle x, y\rangle = \sum_{k=1}^{n} \xi_k \eta_k^* \in \mathbb{C}$$

for each $x \in X$. The dimensions of the factors in the sums can be adjusted by a series of weight coefficients ($w_k \in \mathbb{C}$) by shifting $x_k \to x_k w_k$ for $k = 1, 2, ..., n$. This can be applied in general to a given scalar multiplication:

$$\langle a, b\rangle = \sum_{k=1}^{n} \alpha_k \beta_k^* w_k, \tag{2.34}$$

of arbitrary vectors $a = (\alpha_k)$ and $b = (\beta_k)$ from X. We have further seen that we can interpret scalar multiplications in terms of measurement (starting from quantum mechanics), exchange of action, and information perception.

An example of a dot product in physics is the work (energy) done by a force on a path:

$$W = \vec{F} \cdot \vec{r} = (F_x, F_y, F_z) \cdot (r_x, r_y, r_z) = F_x r_x + F_y r_y + F_z r_z.$$

An example of a norm is the relativistic interval of Minkowski space:

$$ds^2 = -dx^2 - dy^2 - dz^2 + c^2 dt^2$$

in rectangular Cartesian coordinates or in the Schwarzschild metric:

$$ds^2 = -\left(1 - \frac{r_s}{r}\right)^{-1} dr^2 - r^2 d\Omega^2 + \left(1 - \frac{r_s}{r}\right) c^2 dt^2$$

in spherical coordinates. Here $r_s = 2GM/c^2$ Schwarzschild radius, G is the gravitational constant, M the mass of the object, and c the speed of light. Subtractions in these sums give positive values to the square of the interval, because the speed of light is the fastest possible. Weight changes the norm, but when it is just a number it does not change the orthogonality of the vector.

The properties, axioms of the of the *scalar product* (Inner Product) written in Dirac brackets are, for all vectors $x, y, z \in X$ and all scalars $\lambda \in \mathbb{C}$:

1. $\langle x, y\rangle = \langle y, x\rangle^*$;
2. $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$;
3. $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$;
4. if $x \neq 0$, then $\langle x, x\rangle > 0$;

where $z^* = a - ib$ is the complex conjugate of the number $z = a + ib \in \mathbb{C}$, for $a, b \in \mathbb{R}$ and the imaginary unit $i^2 = -1$. Note that axiom 4 defines the norm of a vector, $\|x\|^2 = \langle x, x\rangle$, and that it is assumed to be a positive real number (whenever $x \neq 0$). Therefore, the weights w_k from (2.34) should be consistent with this. Then the previous three conditions are also satisfied. For example:

$$\langle x + y, z\rangle = \sum_{k=1}^{n} (\xi_k + \eta_k)\zeta_k^* w_k = \sum_{k=1}^{n} \xi_k \zeta_k^* w_k + \sum_{k=1}^{n} \eta_k \zeta_k^* w_k = \langle x, z\rangle + \langle y, z\rangle,$$

so the 3rd property is fulfilled.

However, the *scalar product weight* does not have to be a scalar. It can be associated with any vector as an operator. For example, when working with very large and very small numbers in vectors, such as:

$$a = \begin{pmatrix} 1.5 \cdot 10^{-32} \\ 0.8 \cdot 10^{12} \end{pmatrix}, \quad b = \begin{pmatrix} 0.3 \cdot 10^{-32} \\ -1.6 \cdot 10^{12} \end{pmatrix}, \quad c = \begin{pmatrix} -0.3 \cdot 10^{-32} \\ 2.1 \cdot 10^{12} \end{pmatrix},$$

when multiplying, you sometimes need to round the results and lose precision. This can be avoided by introducing a “weight matrix”

$$w = \begin{pmatrix} 10^{32} & 0 \\ 0 & 10^{-12} \end{pmatrix}$$

with which we define a new scalar product. In a simpler way:

$$\langle a, b\rangle_w = \langle a|w|b\rangle = \begin{pmatrix} 1.5 \cdot 10^{-32} & 0.8 \cdot 10^{12} \end{pmatrix} \begin{pmatrix} 10^{32} & 0 \\ 0 & 10^{-12} \end{pmatrix} \begin{pmatrix} 0.3\ 10^{-32} \\ -1.6 \cdot 10^{12} \end{pmatrix} =$$

$$= \begin{pmatrix}1.5 & 0.8\end{pmatrix}\begin{pmatrix}0.3\cdot 10^{-32} \\ -1.6\cdot 10^{12}\end{pmatrix} = 0.45\cdot 10^{-32} - 1.28\cdot 10^{12}.$$

Another way is to redefine each vector with this weight:

$$a' = wa = \begin{pmatrix}10^{32} & 0 \\ 0 & 10^{-12}\end{pmatrix}\begin{pmatrix}1.5\cdot 10^{-32} \\ 0.8\cdot 10^{12}\end{pmatrix} = \begin{pmatrix}1.5 \\ 0.8\end{pmatrix},$$

$$\langle a', b'\rangle = \langle aw, bw\rangle = \langle a|w^\dagger w|b\rangle = 1.5\cdot 0.3 - 0.8\cdot 1.6 = -0.83.$$

The vectors are finally restored to their original state by inverse multiplication $v = w^{-1}v'$. In this case, when the weight matrix is not diagonal, it is better to be hermitian.

Example 101. *Given vectors and weight matrix:*

$$x = \begin{pmatrix}1 \\ 0 \\ i\end{pmatrix}, \quad y = \begin{pmatrix}1 \\ 1-i \\ i\end{pmatrix}, \quad w = \begin{pmatrix}1 & 1-i & i \\ 1+i & 2 & 4 \\ -i & 4 & 3\end{pmatrix}.$$

Find the scalar product $\langle x, y\rangle_w$ and the norms of the vectors $\|x\|_w$, $\|y\|_w$. Compare them with $\langle x, y\rangle$ and $\|x\|$, $\|y\|$.

Solution. Without weights, the squared norms and inner product of the vectors are:

$$\|x\|^2 = \langle x, x\rangle = \begin{pmatrix}1 & 0 & -i\end{pmatrix}\begin{pmatrix}1 \\ 0 \\ i\end{pmatrix} = 2, \quad \|y\|^2 = 4, \quad \langle x, y\rangle = 0.$$

The matrix is Hermitian, equal to its conjugate-transpose $w^\dagger = w$, so we calculate:

$$\|x\|_w^2 = \langle x|w^\dagger w|x\rangle = \begin{pmatrix}1 & 0 & -i\end{pmatrix}\begin{pmatrix}1 & 1-i & i \\ 1+i & 2 & 4 \\ -i & 4 & 3\end{pmatrix}^2\begin{pmatrix}1 \\ 0 \\ i\end{pmatrix} =$$

$$= \begin{pmatrix}1 & 0 & -i\end{pmatrix}\begin{pmatrix}4 & 3+i & 4 \\ 3-i & 22 & 19+i \\ 4 & 19-i & 26\end{pmatrix}\begin{pmatrix}1 \\ 0 \\ i\end{pmatrix} = \begin{pmatrix}4 & 4-18i & 4-26i\end{pmatrix}\begin{pmatrix}1 \\ 0 \\ i\end{pmatrix} = 30,$$

$$\|y\|_w^2 = \langle y|w^\dagger w|y\rangle = \begin{pmatrix}1 & 1+i & i\end{pmatrix}\begin{pmatrix}4 & 3+i & 4 \\ 3-i & 22 & 19+i \\ 4 & 19-i & 26\end{pmatrix}\begin{pmatrix}1 \\ 1-i \\ i\end{pmatrix} =$$

$$= \begin{pmatrix}8-2i & 24+4i & 22-6i\end{pmatrix}\begin{pmatrix}1 \\ 1-i \\ i\end{pmatrix} = 42,$$

$$\langle x, y\rangle_w = \langle x|w^\dagger w|y\rangle = \begin{pmatrix}1 & 0 & -i\end{pmatrix}\begin{pmatrix}4 & 3+i & 4 \\ 3-i & 22 & 19+i \\ 4 & 19-i & 26\end{pmatrix}\begin{pmatrix}1 \\ 1-i \\ i\end{pmatrix} =$$

$$= \begin{pmatrix} 4-4i & 2-18i & 4-26i \end{pmatrix} \begin{pmatrix} 1 \\ 1-i \\ i \end{pmatrix} = 14-20i.$$

Orthogonal vectors with matrix weight are no longer. □

The orthogonality of Bessel functions of the first kind (example 100) with weight $w(x) = x$ has been proven, and here is another example from the literature[8].

Example 102. *The functions:*

$$f_0(x) = 1, \quad f_1(x) = 2x, \quad f_2(x) = 4x^2 - 1, \quad f_3(x) = 8x^3 - 4x,$$

$$f_4(x) = 16x^4 - 12x^2 + 1, \quad f_5(x) = 32x^5 - 32x^3 + 6x$$

are orthogonal to $[-1, 1]$ *with weight* $w(x) = \sqrt{1-x^2}$.

Explanation. We check orthogonality by integrating:

$$\int_{-1}^{1} f_0(x) f_1(x) w(x)\, dx = -\int_{-1}^{1} \sqrt{1-x^2}\, d(1-x^2) = -\frac{(1-x^2)^{3/2}}{3/2}\bigg|_{-1}^{1} = 0,$$

and further, $\int_{-1}^{1} f_i(x) f_j(x) w(x)\, dx$, for each pair of indices $i, j \in \{0, 1, 2, 3, 4, 5\}$. □

Analogous to the expansion of the function $f(x)$ in the (general Fourier) order (2.27) of "ordinary"orthogonal functions, it is also possible to expand it into the order of "weighted"orthogonal functions. Do it yourself with the aforementioned Bessel functions, and let's look at a similar expansion in terms of weighted orthogonal functions from the previous example. The form is the same:

$$g(x) = \sum_k a_k f_k(x), \tag{2.35}$$

but the coefficients are "weighted":

$$a_n = \frac{\langle g, f_k \rangle}{\langle f_k, f_k \rangle} = \frac{\int_a^b g(x) f_k^*(x) w(x)\, dx}{\int_a^b |f_k(x)|^2 w(x)\, dx}.$$

The calculation is therefore a little more extensive.

Example 103. *Let's expand the function* $g(x) = x^5 - 1$ *into a sum of weighted orthogonal polynomials from the previous, 102nd example.*

Solution. First, let's calculate the coefficients:

$$a_0 = -1, \quad a_1 = \frac{5}{32}, \quad a_2 = 0, \quad a_3 = \frac{1}{8}, \quad a_4 = 0, \quad a_5 = \frac{1}{32},$$

[8] Ryan C. Daileda: Partial Differential Equations, Trinity University, 2014.

which we need to check. Say:

$$a_1 = \frac{\langle x^5 - 1, 2x\rangle}{\langle 2x, 2x\rangle} = \frac{\int_{-1}^{1} 2x(x^5-1)\sqrt{1-x^2}\,dx}{\int_{-1}^{1} 4x^2\sqrt{1-x^2}\,dx} = \frac{5\pi/64}{\pi/2} = \frac{5}{32}.$$

Hence:

$$g(x) = -f_0(x) + \frac{5}{32}f_1(x) + \frac{1}{8}f_3(x) + \frac{1}{32}f_5(x) =$$
$$= -1 + \frac{5}{32}(2x) + \frac{1}{8}(8x^3 - 4x) + \frac{1}{32}(32x^5 - 32x^3 + 6x)$$
$$= -1 + \frac{5}{16}x + x^3 - \frac{1}{2}x + x^5 - x^3 + \frac{3}{16}x$$
$$= x^5 - 1,$$

and this is correct. □

The weight function w allows for the normalization of the polynomial f_k. However, this orthogonality is not necessary for a function like g from similar examples to be represented as a linear combination $g = a_0 f_0 + a_1 f_1 + \dots + a_n f_n$. The linear independence of the factors f_k is sufficient for these vectors to span the vector space g. If they succeed, that notation is unique.

Namely, if there were another record $g = b_0 f_0 + b_1 f_1 + \dots + b_n f_n$, by subtraction we find $(a_0 - b_0)f_0 + (a_1 - b_1)f_1 + \dots + (a_n - b_n)f_n = 0$, so due to the linear independence of the basis vectors f_k we have $a_k - b_k = 0$ for each $k = 0, 1, \dots, n$. Therefore, $b_0 = a_0$, $b_1 = a_1$, ..., $b_n = a_n$. Possible orthogonality is only a "spice", that is, a way for the vector g to see the components of f_k as independent "observables". Thus g becomes a "subject" of the space as defined by the basis vectors f_k.

Example 104. *Let's represent the function $g(x) = x^5 - 1$ as a linear combination of the polynomials of example 102.*

Solution. Into the required equation $g = b_0 f_0 + b_1 f_1 + \dots + b_n f_n$ we insert the given (which can be some other given) polynomials:

$$x^5 - 1 = b_0 \cdot 1 + b_1 \cdot 2x + b_2(4x^2 - 1) + b_3(8x^3 - 4x) +$$
$$+ b_4(16x^4 - 12x^2 + 1) + b_5(32x^5 - 32x^3 + 6x) =$$
$$= (b_0 - b_2 + b_4) + (2b_1 - 4b_3 + 6b_5)x + (4b_2 - 12b_4)x^2 +$$
$$+ (8b_3 - 32b_5)x^3 + 16b_4x^4 + 32b_5x^5.$$

We equate the coefficients of the polynomials on the left and right of the equality:

$$b_0 - b_2 + b_4 = -1, \quad 2b_1 - 4b_3 + 6b_5 = 0, \quad 4b_2 - 12b_4 = 0,$$
$$8b_3 - 32b_5 = 0, \quad 16b_4 = 0, \quad 32b_5 = 1,$$
$$b_0 = -1, \quad b_1 = \frac{5}{16}, \quad b_2 = 0, \quad b_3 = \frac{1}{8}, \quad b_4 = 0, \quad b_5 = \frac{1}{32}.$$

This result and from Example 103 are equal expansions of the function g. □

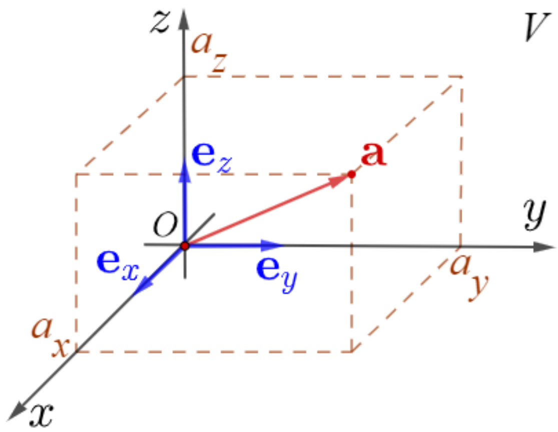
The diversity of the weights $w(x)$ that leads to different families of orthogonal polynomials (such as Legendre, Chebyshev, Hermite, Lager, or other, simpler ones) also affects the convergence conditions, zeros, and similar forms of the properties of the polynomials, and interpreted, such adjust the ways of perceiving *information coupling* of the subject and the object.

The subject is unique, as are arrays like this b_k which often have more coefficients $k = 0, 1, 2, \ldots$, compared to the perhaps many available observables f_k of the environment. The weight w can then simply be the probability of success in measuring the observable (in the case of physical interactions), or the importance of that observable in the life of the individual. These differences are reflected in the better vision of some birds, the auditory orientation of bats, the excellent sense of smell of bears, or the cognitive perceptions of humans.

2.4.9 Fermions and bosons

The uniqueness of representing a vector by a linear combination of basis vectors can be seen in the following figure on the right. It is a 3-dimensional Cartesian rectangular coordinate system $(Oxyz)$ with *orths*, orthogonal unit vectors $\mathbf{e}_x$, $\mathbf{e}_y$ and $\mathbf{e}_z$ abscissas, ordinates and applicates. It defines a vector space V.

Vector $\mathbf{a} = a_x\mathbf{e}_x + a_y\mathbf{e}_y + a_z\mathbf{e}_z$ extends along a_x units of abscissa length, a_y units of ordinate length, and a_z units of applicate. It is just one single point, the vertex of the parallelepiped in the figure on the right. And such uniqueness has far-reaching consequences that are the topic of our discussion here.



Its application (Riesz's theorem) in functionals $f : V \to \Phi$ has often been mentioned in the text, as the existence of a unique vector $\mathbf{b} \in V$ which, by scalar multiplication with an arbitrary vector $\mathbf{a} \in V$, gives the value of the given functional, $f(\mathbf{a}) = \langle \mathbf{a}, \mathbf{b} \rangle$.

The further purpose, now both sentences, is in information of perception. The coupling of information of a subject and an object that can communicate, or interact, or measure each other, is the interpretation of the functional f. The specific subject is then the mentioned vector $\mathbf{b}$, and what it can perceive around itself are the vectors $\mathbf{a}$.

The spread of this non-repeatability as an abstract and highly repeatable model in various occasions is the other side of the coin. There is no practice without truths that we have, or perhaps will never discover, and which are easily multiplied everywhere (Skolem's theorem). We saw this as the indeterminacy in solving a system of linear equations (example 87), when there are not enough basis vectors

to define a given vector and functional. Paraphrasing, individuals are unique, but their representations are not.

We recognize the theoretical as that which is repeatable in the concrete. For example, when we toss two coins and write down for each outcome "tails" or "heads", we have four possibilities {tt, th, ht, hh}. We also have this when we toss the same coin twice. To see if there is a natural difference between the "th" and "ht" outcomes, we can repeat this experiment many times and see if, say, "tt" is a quarter, or a third, of all the outcomes. Since this is approximately a quarter of all these double outcomes, this indicates the equality of each of the four listed, that is, the natural distinction of the order of the "th" and "ht" outcomes, regardless of whether it is a single coin tossed twice, or two coins tossed at once.

I cite this example only as a difficulty or possibility of experimentally distinguishing a "subject" from an "abstraction". The second toss and the repetition of the trial with the same probabilities, according to this, is a theory. When we repeat a binary event with the probability of the desired outcome $p \in (0,1)$ and the probability of the other $q = 1 - p$ expecting the desired one to appear in the nth step, the probability of such an outcome has an exponential distribution $\Pr(n) = q^{n-1}p = e^{-\lambda n}$ for $\lambda = \ln 2$. Indeed:

$$\Pr(1) + \Pr(2) + \Pr(3) + ... = p + qp + q^2p + ... = \lim_{n\to\infty} \frac{1-q^n}{1-q}p = \frac{1}{1-q}p = 1,$$

and we have a probability distribution. Then:

$$e^{-\lambda} + e^{-2\lambda} + e^{-3\lambda} + ... = 1,$$

$$e^{-\lambda}(1 + e^{-\lambda} + e^{-2\lambda} + ...) = 1,$$

$$e^{-\lambda}\frac{1}{1-e^{-\lambda}} = 1,$$

and hence $e^{-\lambda} = 1 - e^{-\lambda}$ and $2e^{-\lambda} = 1$, so $2 = e^{-\lambda}$ and $\ln 2 = \lambda$. That's the theory.

However, there is also practice, because (Extremes) the exponential distribution has the maximum information of all those with a given final expectation. Since nature spontaneously avoids greater information (the principle of minimalism), in practice there is also "memory"of past outcomes and their influence on the next, such as corrupting the exponential distribution. For example, theoretically, germs multiply exponentially, but practically, they kill future victims by thinning the environment, or surround themselves with immune ones and the exponential spread of the disease slows down.

Also, each transformed state of an energy system can be the matter of a concrete subject, but the law of conservation of quantity is a theoretical property for them. In the world of particles, fermions are "subjects", and bosons are "concepts". Conditionally speaking, hence the quotation marks. In the case of the former, Pauli's (1925) exclusion principle states that electrons in the same atom cannot have identical key values. No two electrons in the same orbital can have all four quantum numbers: n (principal), ℓ (azimuthal), m_ℓ (magnetic), and m_s (spin).

This principle was used to understand and build the Periodic table of elements, for which Pauli received the Nobel Prize in Physics in 1945. More broadly, this is a detail of the "unrepeatability of subjects", because no two atoms can be in the same place, i.e. each atom is unique in relation to its environment.

However, it is not always correct to try to treat *composite systems* as elementary. The protons and neutrons in the nucleus of an atom are composite particles, because they are made up of quarks. A subatomic particle that is made up of two or more elementary particles is also composite, but the electrons orbiting the nucleus are not, they are *elementary particles*. Fundamental, elementary particles are those that have no subparts and cannot be divided. All leptons are elementary particles, but the neutron and the proton are not and instead are made up of three fundamental particles, called quarks, which have been observed through high-energy physics experiments.

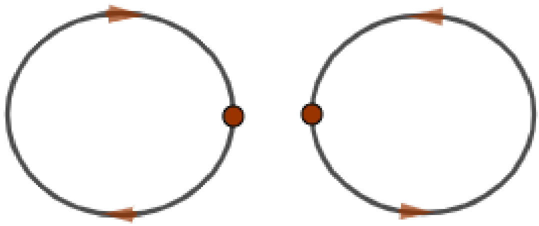
Elementary particles are divided into bosons and fermions. We distinguish them by the symmetry of the wave function when exchanging particles. The state of the particles is represented by $\psi(r_1, r_2, ..., r_n)$, where rk is the position of the k-th particle. Not differentiating the identical means exchanging the wave function by a factor P, i.e. $\psi(..., r_i, ..., r_j, ...) = P\psi(..., r_j, ..., r_i, ...)$. Exchanging two particles twice returns the system to the original situation, so that $P^2 = 1$, and hence $P = \pm 1$. More generally, these factors (then there are more of them) are *Pauli matrices* (2.4) whose square is the identity matrix.

The choice of sign, or matrix P, is closely related to the spin, the quantum number m_s of the particles. Namely, particles with integer spin have $P = +1$ and are called bosons, while those with half-integer spin have $P = -1$ and are called fermions. This can be expressed by writing $P = e^{i\phi}$, where the phase of the ϕ wave function takes on the values $k\pi$ ($k \in \mathbb{Z}$) when exchanging particles. Fermions and bosons exhibit quite different behaviors. For example, at very low temperatures bosons undergo Bose-Einstein condensation and become a superfluid.

A Bose–Einstein condensate is a gas of extremely low density and very low temperature, which reduces its kinetic energy. At absolute zero, particles have no energy to move and occupy only the lowest energy quantum state. Unlike fermions, which must satisfy the Pauli exclusion principle, bosons will all be in exactly the same quantum state at absolute zero, where they are treated as absolutely indistinguishable identical particles.

Electrons are fermions with spin $\pm\frac{1}{2}$. Our three-dimensional (3-dim) world strictly distinguishes between fermions and bosons, but in the two-dimensional (2-dim) world, *flux binding* can occur. A fermion can be transformed into a boson (or vice versa) by attaching a fictitious magnetic flux to it that changes the phases of its wave function (Fermions become bosons).

Namely, the exchange of two particles in two dimensions, the left one clockwise and the right one counterclockwise (in the figure on the left) will produce a conserved *winding number* in two dimensions. This winding number (Winding number) of a closed curve in the plane around a given point is the integer number of counterclockwise orbits around the given point.



The clockwise and counterclockwise exchange operations are different in two dimensions. They are not different in three dimensions, because one can be changed into the other by changing the path. The way the phase of the wave function can be changed when exchanging particles is by adding a fictitious flux to each particle.

Glava 3

External product

The vector or outer product of two contravariant u^i and v^i (or covariant u_i and v_i) vectors in a three-dimensional space 3-dim using the relative tensor e_{ijk} is defined by

$$w_i = e_{ijk} u^j v^k \tag{3.1}$$

or, more precisely written:

$$w_1 = u^2 v^3 - u^3 v^2, \quad w_2 = u^3 v^1 - u^1 v^3, \quad w_3 = u^1 v^2 - u^2 v^1.$$

That is a *commutator*. The commutator in Euclidean space (of the norm $\|x\|_2 = \|x\|$) is equal to the area spanned by the vectors $\vec{u} = (u^1, u^2, u^3)$ and $\vec{v} = (v^1, v^2, v^3)$. Assuming that $\vec{w} = (w_1, w_2, w_3)$ then:

$$\|w\| = [\vec{u}, \vec{v}] = \|u\| \|v\| \sin \angle(\vec{u}, \vec{v}).$$

This is the intensity of the vector product of the vectors $\vec{w} = \vec{u} \times \vec{v}$. Its direction is perpendicular to the plane of the vectors $\vec{u}$ and $\vec{v}$, and the direction is determined by the "right-hand rule." The special story about commutators here is their connection of the subject with external permanent central forces.

3.1 Maxwell's equations

Maxwell's equations (1864) are the foundations of the theory of electromagnetism. Some of them were known before and are still called Gauss's law, Gauss's law of magnetism, Faraday's law and Ampere's law (with Maxwell's correction), but Maxwell was the first to unify them and supplement them with his discoveries.

1. $\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}$;
2. $\nabla \cdot \mathbf{B} = 0$;
3. $\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$;
4. $\nabla \times \mathbf{B} = \mu_0 \cdot \mathbf{J} + \mu_0 \cdot \epsilon_0 \cdot \frac{\partial \mathbf{E}}{\partial t}$;

These are for vacuum in the SI system. The vectors $\mathbf{E}$, $\mathbf{B}$ and $\mathbf{J}$ are the electric, magnetic field and electric current density, respectively, and $\mu_0\epsilon_0 = 1/c^2$ is the vacuum permeability multiplied by the vacuum dielectric constant, with c the speed of light on the right-hand side of the equation.

For the electric and magnetic fields, we use the notations $\mathbf{D} = \epsilon_0\mathbf{E}$ and for the magnetic flux density $\mathbf{B} = \mu_0\mathbf{H}$, so the first equation, called *Gauss's law*, can be written $\nabla\cdot\mathbf{D} = \rho$, and the second, called Gauss's law of magnetism (magnetic monopoles do not exist), can be written $\nabla\cdot\mathbf{H} = 0$. The third is *Faraday's law*, and the fourth is *Ampere's law* $\nabla\times\mathbf{H} = \partial_t\mathbf{D}+\mathbf{J}$. Otherwise $\mathbf{J} = \sigma\mathbf{E}$ is *Ohm's law* with material conductivity σ, so this equation becomes $\nabla\times\mathbf{H} = \partial_t\mathbf{D} + \sigma\mathbf{E}$.

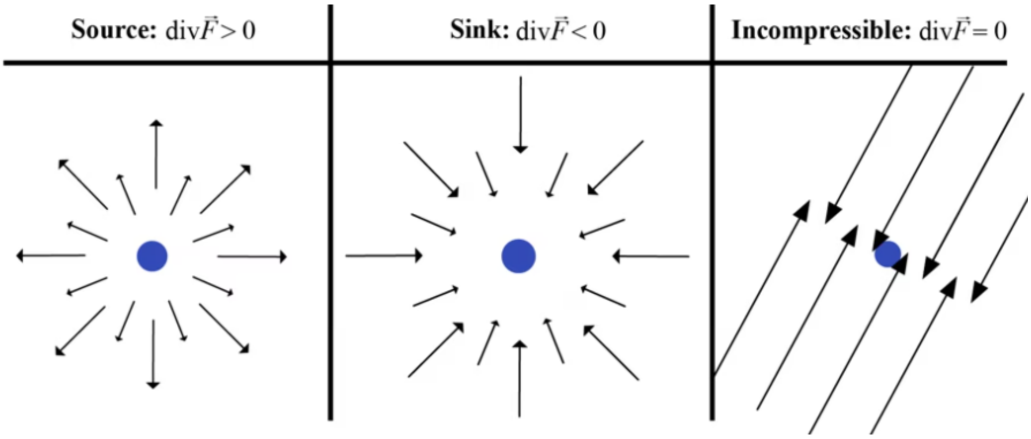
The Nabla operator is divergence, $\nabla\cdot = \text{div}$, in the first two equations:

$$\nabla\cdot\mathbf{E} = \text{div }\mathbf{E} = (\mathbf{e}_x\frac{\partial}{\partial x} + \mathbf{e}_y\frac{\partial}{\partial y} + \mathbf{e}_z\frac{\partial}{\partial z})\cdot(\mathbf{e}_x\frac{\partial}{\partial x} + \mathbf{e}_y\frac{\partial}{\partial y} + \mathbf{e}_z E_z) = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}.$$

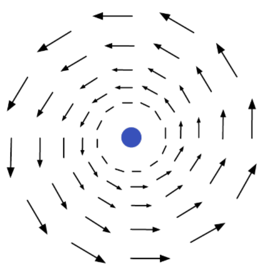
The Nabla operator is rotor, $\nabla\times = \text{rot}$, in the last two equations:

$$\nabla\times\mathbf{E} = \text{rot }\mathbf{E} = \begin{vmatrix} \mathbf{e}_x & \mathbf{e}_y & \mathbf{e}_z \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ E_x & E_y & E_z \end{vmatrix}.$$

In the figure (calcworkshop) is the divergence of the force $\mathbf{F}$: source, sink and neutral.



The rotor represents a circle, as in the following figure on the right. Around the direction of the vector, say around a conductor of electric current, a magnetic field rotates. The basic rotation theorem is Stokes. It connects the curve integral around a simple closed curve C and the double integral over the region bounded by that curve:



$$\int_C \mathbf{F}\cdot\mathbf{r} = \iint_S \nabla\times\mathbf{F}\cdot d\mathbf{S}.$$

In other words, the total curvature of the region is equal to the curvature at its boundary. Instead of a double integral, a circle is drawn on the

integral symbol to indicate the closure of the curve C, so Faraday's and Ampere's laws (3 and 4. Maxwell's) we write:

$$\oint_{\partial\Sigma} \mathbf{E}\cdot d\mathbf{l} = \iint_{\Sigma} \nabla\times\mathbf{E}\cdot d\mathbf{S}, \quad \oint_{\partial\Sigma} \mathbf{B}\cdot d\mathbf{l} = \iint_{\Sigma} \nabla\times\mathbf{B}\cdot d\mathbf{S}.$$

The dielectric constant of vacuum $\epsilon_0 = 8.854\times 10^{-12}\ \mathrm{F}\cdot\mathrm{m}^{-1}$, Farad per meter. The farad is the SI unit of electrical capacitance named after Michael Faraday, where one coulomb of charge produces a potential difference of one volt (C/V). It is also called vacuum permittivity.

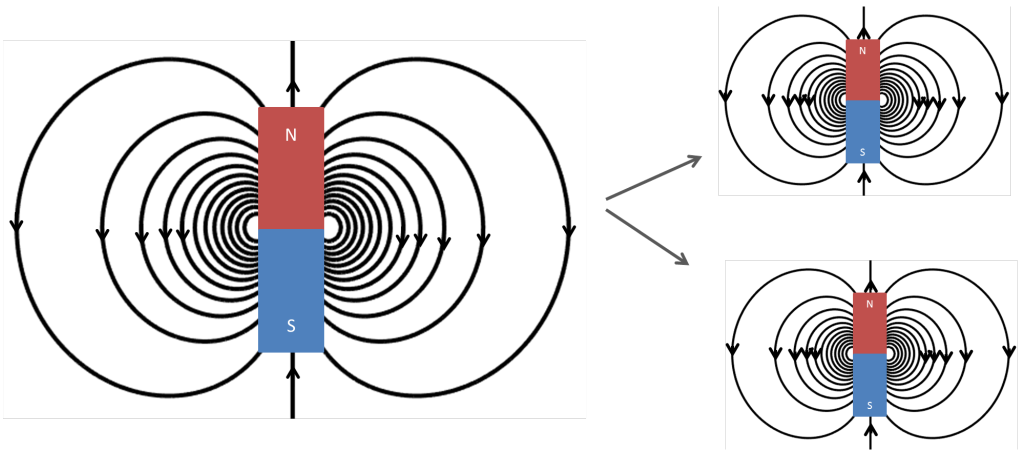
Example 105. *What is the electric current through the surface of a sphere containing 12 protons and 8 electrons?*

Solution. Gauss's law (Maxwell's 1st law) for electric fields relates the electric flux through a closed surface to the charge enclosed by that surface:

$$\Phi_E = \oint_S \nabla\cdot\mathbf{E}\ dS = \frac{\sum_i q_i}{\epsilon_0} = \frac{(12-8)(1.6\cdot 10^{-19}\ \mathrm{C})}{8.854\times 19^{-12}\ \mathrm{C/Vm}} = 7.228\times 10^{-8}\ \mathrm{Vm}.$$

The size of the sphere is irrelevant. □

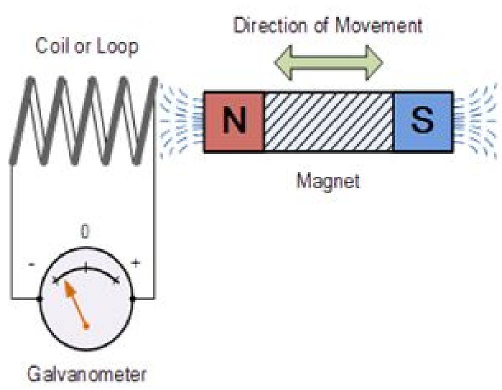
These ideas about flux are very old in relation to the discovery of electrons, protons and the true nature of charge. Fortunately, these formulas were made so general that they can be applied to any fluid flow for which conservation laws would apply and, therefore, they are easily transferred to the corresponding emissions of information (quantities of uncertainty). It seems like an extreme case, but here there may also be information that "radiates"from the present to the future, like a 4-dim sphere, with which we go from our past and through the present to the future.



When a bar magnet is cut in two, we get two magnets. This is explained by Maxwell's 2nd law, or Gauss's law of magnetism. The idea itself, about the non-existence of magnetic monopoles, was established long ago (Petrus Peregrinus de

Maricourt, 1269). However, Gauss's work was also greatly influenced by Gilbert's book De Magnete (1600). Experiments on only one pole of a magnet were reproduced by Faraday (1800) and finally included in the field equations by Maxwell (1864).

A permanent magnetic field originates at the north pole (N) and disappears at the south pole (S). The interesting question is then where are all the other models of such (theoretical) behavior, but I will leave the answer for another occasion.



To interpret Faraday's law, or Maxwell's 3rd equation, we first familiarize ourselves with an experiment like the one shown in the figure on the left. We have a coil connected to a galvanometer and a separate bar magnet. The coil has no current source, no battery connected, and when the magnet is stationary, no current flows through the coil. However, when the bar magnet is moved towards the coil, the galvanometer produces a deflection. The movement of the magnet induces a current.

Changes in the magnetic flux in a circuit produce an induced ElectroMotive Force. The EMF lasts only as long as the flux changes, and its magnitude is directly proportional to the rate of change of the magnetic flux associated with the electrical circuit.

Example 106. *The magnetic field is given as a function of time by the vector*

$$\mathbf{B}(t) = B_0 \cos(kz - \omega t)\mathbf{e}_y.$$

(a) Find the rotor induced electric field.

(b) If $E_z = 0$, what is E_x?

Solution. (a) By Faraday's law, the rotor electric field is the negative derivative of the vector magnetic field with respect to time. So:

$$\text{rot } \mathbf{E} = \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} = -\frac{\partial(B_0 \cos(kz - \omega t))}{\partial t},$$

$$\text{rot } \mathbf{E} = -\omega B_0 \sin(kz - \omega t)\mathbf{e}_y.$$

(b) Writing out the components of the rotor gives:

$$\left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}\right)\mathbf{e}_x + \left(\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x}\right)\mathbf{e}_y + \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}\right)\mathbf{e}_z = -\omega B_0 \sin(kz - \omega t)\mathbf{e}_y.$$

Equating the $\mathbf{e}_y$ components and setting E_z to zero, we get:

$$\frac{\partial E_x}{\partial z} = -\omega B_0 \sin(kz - \omega t),$$

$$E_x = \int -\omega B_0 \sin(kz - \omega t)\, dz = \frac{\omega}{k} B_0 \cos(kz - \omega t).$$

This is the result within the constant of integration. □

Maxwell's fourth equation (revised Ampere's law) is elegantly expressed in integral form:

$$\oint \mathbf{B} \cdot d\mathbf{l} = \mu_0 \epsilon_0 \frac{d}{dt} \int \mathbf{E} \cdot d\mathbf{a} + \mu_0 \int \mathbf{J} \cdot d\mathbf{a}$$

where $\mathbf{B}$ is the magnetic field vector, μ_0 is the permittivity (dielectric constant) of free space, $\mathbf{E}$ is the electric field vector, $d\mathbf{a}$ is the area differential vector, and the vector $\mathbf{J}$ is the current density. A typical analogy is the flow of water in a river. When the width of the river (of the changing electric field) is narrowed, the flow of water is accelerated (similar to an electric current).

Maxwell's fourth equation essentially states that the magnetic field around a closed loop is equal to the sum of the electric current flowing through the loop and the rate of change of the electric field multiplied by a constant. This is what we mean by the radiation of electromagnetic waves, such as light. When the electric field changes rapidly, as in an alternating current circuit, the displacement current becomes significant and leads to the propagation of electromagnetic waves at the speed of light $c = 1/\sqrt{\mu_0 \epsilon_0}$.

Example 107. *Determine the magnetic field of a closed wire with a radius of 0.5 m (meters) if a current of 2 A (amperes) flows through it.*

Solution. The radius of the conductor is $r = 0.5$ m, the current is $I = 2$ A, and the permittivity of vacuum is $\mu_0 = 4\pi \times 10^{-7}$ N/A^2 (newtons per ampere squared), so we calculate the magnetic field strength:

$$B = \frac{\mu I}{2\pi r} = \frac{4\pi \times 10^{-7} \cdot 2}{2\pi \cdot 0.5} = 8 \cdot 10^{-7} \times \text{ N m}^{-1}$$

newtons per meter (surface tension). □

The comment with the 105th example for the 1st Maxwell equation was also the "radiation of 4-dim spheres" of information of the present that goes from the past to the future. The one with the 4th Maxwell equation and the observation that the mentioned "sphere" grows (the universe is expanding), like the decreasing speed of water flow in a wider river, leads to the conclusion that the speed of "radiation" decreases. By the principle of minimalism, the present should go towards a more certain future, and therefore a lower density of options, rarer events and a slower flow of time. The third way (of my) information theory towards the same is the attraction of a slower flow of time.

I assume, for now, that fermions (elementary particles of matter) could be changing into elementary bosons (particles of space) a little more often than the other way around, and that, along with perhaps something else, this could be the reason for the expansion of the universe. However, the previous thesis and its three stated reasons do not require this. It is not at all necessary to know why and how fast the universe is expanding to conclude about the trend towards certainty. The (perhaps deceptive) idea has simply taken hold of itself, that distant past spaces only seem to be getting bigger to us, because our units of length are smaller and time longer.

3.1.1 Retrofitting

Apparently, one of the obstacles to applying the flux from Maxwell's equations and information is the one-way flow of events, from the past to the future and never vice versa. The flows described by the laws of electromagnetism in their 3-dimensional space can take various directions, while in 4-dimensional space-time, it is believed, there is no way for the future to affect the past. But is this really so?

Back in the days of the founder of quantum mechanics (Heisenberg, 1924), it was observed that "the path of an electron is created by measurement." There are such elementary particles that their entire communication is that one possible interaction that can drastically change them. Below is (my) computer science description of the possible contact of an elementary particle with a measuring device. Keep in mind that in this theory, "reality" is a relative phenomenon. It exists only for those subjects who could communicate with it, even indirectly.

An elementary particle-wave communicates by interacting with its environment. It contacts another object by transferring its information to it, which means some "quantity of uncertainty" (specifically a defined measure), after which its certainty increases in the same "reality" in which it participates with its act. Unlike large bodies, say, the Moon, for which it is unimportant whether any of the many observers is observing it at a given moment, each individual act of collision is a drastic emission of its "volume of options" for a small particle.

By losing options, by reducing uncertainty, its certainty increases, which also includes objectively clearer its previous trajectory. Objectivity here is in the sense of the stated reality, therefore: 1. the possibility of observation, 2. some constancy, 3. truthfulness. The action on the past is thus reduced to micro-phenomena created by measuring in the present, which further defines something that was, that is, which reduces the number of options of what could have been. Then, retroactively, the newly created situation from the past will narrow the current options of the present and to that extent (I am talking about a broadly negligible action) clarify the future.

The present moves away more and more certainly, leaving a trace of the past behind it, which, in turn, just compensates for the loss of the current information of the universe so much that its total amount remains unchanged. I wrote about this "illusion of duration" earlier in the framework of the Markov chain, which ends with a "black box" with some distant number of links. What we see in the deep past, at the end of this long series of information carriers, is a kind of "illusion." However, its persuasiveness is such that it is difficult to refute it by astronomical observation or by experiment, and in general by logic of lower truth than mathematical.

Let's see if this has (or doesn't have) an analogy with regular assignments from theoretical physics classes, which was an elective subject for me (along with quantum mechanics) at my mathematics studies in Belgrade in the early 80s.

Example 108. *Let us find the electric field of a uniformly (volumetrically) charged sphere of radius R, if its total charge is Q.*

Solution. Let r be the distance of a given point from O, the center of the sphere.

The volume of a sphere of radius r is $\frac{4}{3}r^3\pi$. The given charge density is:

$$\rho(r) = \begin{cases} \frac{3Q}{4\pi R^3}, & r \le R \\ 0, & r > R \end{cases}$$

as a result of which the regions inside and outside the sphere must be considered separately. Due to spherical symmetry, the electric field at all points is radial, and for the surface A of the sphere, it is

$$\oint_S \mathbf{E} \cdot d\mathbf{A} = 4\pi r^2 E(r)$$

due to the collinearity (same direction) of the vectors $\mathbf{E}$ and $d\mathbf{A}$. Hence:

$$4\pi r^2 E(r) = \frac{1}{\epsilon_0}\int_0^r \rho \cdot 4\pi r^2\, dr = \frac{1}{\epsilon_0} \cdot \rho \cdot \frac{4\pi r^3}{3} = \frac{1}{\epsilon_0}\frac{Qr^3}{R^3}, \quad r \le R,$$

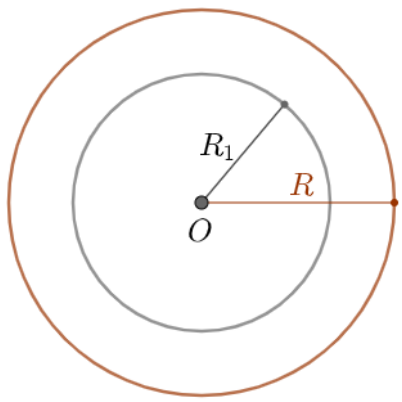
$$4\pi r^2 E(r) = \frac{1}{\epsilon_0}\int_0^R \rho \cdot 4\pi r^2\, dr = \frac{1}{\epsilon_0} \cdot Q, \quad r > R.$$

The final result can be written in the form:

$$E(r) = \begin{cases} \frac{1}{4\pi\epsilon_0}\frac{Qr}{R^3}, & r \le R, \\ \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2}, & r > R. \end{cases}$$

Inside the ball, the electric field intensity increases proportionally with the distance from its center, and the external field is identical to the field of a point charge concentrated in the center. □

The analogy of this electric field intensity and the "information emission"of the 4-dim universe is with the past, which has its beginning in the Big Bang (the center of the ball) and becomes proportionally clearer as it approaches the present (the surface of the ball).



If the ball were empty (hollow) up to the radius R_1, and uniformly filled from R_1 to the end (up to R), as shown on the right, then the previous situation would occur again, with an analogous solution: $E(r) = 0$ when $r < R_1$, but $E(r) = \frac{1}{4\pi\epsilon_0}\frac{Q(r-R_1)}{R^3}$ if $R_1 \le r \le R$ and $E(r) = \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2}$ for $r > R$. The electric field behaves as if there were no inner sphere, up to the radius R_1. It is as if it were a single point from which a sphere from R_1 to R begins, with the previous solution including an external field ($r > R$) that is identical to a point charge from the center O.

Similarly, what was before the Big Bang, about 13.8 billion years ago, is as if it never existed, even if it đid"exist. Another issue is the neglectability of the above possible effects of the present on the past.

This is the “problem” of time asymmetry. It emerges from principled minimalism and spontaneously guides events in nature that, it seems, do not exist in the mathematical equations of physics. But that is not quite so either. Let’s look at this on those same Maxwell equations, there in differential form and now in integral form so that I don’t simply repeat them:

1. $\oint \mathbf{E}\cdot d\mathbf{A} = \frac{Q}{\epsilon_0}$;
2. $\oint \mathbf{B}\cdot d\mathbf{A} = 0$;
3. $\oint \mathbf{E}\cdot d\mathbf{l} = -\frac{\partial \mathbf{B}}{\partial t}$;
4. $\oint \mathbf{B}\cdot d\mathbf{l} = \mu_0\epsilon_0\frac{\partial \mathbf{E}}{\partial t} + \mu_0\mathbf{J}$;

The electric $\mathbf{E}$ and magnetic $\mathbf{B}$ field vectors have the direction of the vector $d\mathbf{A}$ of the area increment in which the total charge $Q = \sum_i q_i$ is enclosed. The first two equations already show the asymmetry of nature in the possibility of having an electric field within a given area without such a magnetic field. There are separate particles of electric charge, say negative electrons and positive protons, but not such magnetic ones. Physical reality does not know magnetic monopoles.

The third and fourth equations speak of the electric and magnetic fields through a closed line of increment $d\mathbf{l}$, a circular conductor or coil. The third says that a change in the magnetic field gives an opposite direction to the change in the electric field along the conductor, and the fourth that changes in the electric field, now multiplied by the constants $\mu_0\epsilon_0 = 1/c^2$ plus the current multiplied by the permeability of the medium μ_0, give changes in the magnetic field. Here again, because of the addition of the current $\mu_0\mathbf{J}$, and also because of the factor $1/c^2$ to the change in the electric field, there is no symmetry between the third and fourth equations.

Looking at the left side of the 4th equation and only the first of the two sums, that is, in the absence of additional currents, we see a wave of electromagnetic radiation moving at a constant speed of about $c = 300\,000$ km/s. This is the speed of light, and it is Maxwell’s merit that he saw and understood it as electromagnetic radiation. Einstein later, in his explanations of the special theory of relativity, pointed out that this equation also speaks of the independence of the speed of light from the speed of the source.

The asymmetry of electric and magnetic fields once led me to think about dividing "existence"into real and unreal and about accepting mathematical untruths as this second form of existence. But why would something that is not permanent and tends to arise "out of nothing"and disappear "into nothing"be obliged to appear at all? – was once a favorite riddle of mine. Of course, we now know the answer, due to the equivalence of the worlds of truth and falsehood (1.3 Truth). The latter can disappear and arise in any way, provided only that there are as many of them as the former and that the more probable ones have priority.

After this “phenomenal” discovery (so I thought at the time), electromagnetism was no longer important to me for a long time. I will only mention that it was from there that I first understood the “speed” of the creation of reality at the speed of light.

3.1.2 Hall effect

Hall effect (Edwin Herbert Hall, 1879) is the phenomenon of a transverse voltage within a solid material through which a current is passed and is located in an external transverse magnetic field. When negative currents (today electrons) are connected by a conductor, positive Hall side currents (now we know that they are protons) would be “holes” as additional charge carriers. Lord Kelvin, who was one of the most famous scientists at the time, compared Hall’s discovery to the greatest because little was known about electricity at that time. Only 10 years later was the electron discovered.

If we denote by $\mathbf{F}$ the Lorentz force vector, Q the amount of charge under the action of an external magnetic field vector $\mathbf{B}$ and by $\mathbf{v}$ the velocity vector of the charge carrier, then

$$\mathbf{F} = Q \cdot \mathbf{v} \times \mathbf{B}. \tag{3.2}$$

The formula describes the (Laplace) magnetic force on a conductor, the electromotive force in a wire loop through a magnetic field (by Faraday induction), and the force on a moving charged particle. The law was in Maxwell’s work (1865), but Lorentz (1895) came to a complete derivation and contribution of the electric force a few years after Heaviside’s discovery of the contribution of the magnetic force.

Example 109. *Let’s find the equations of motion of a particle of mass m and charge $-e$ in a magnetic field $\mathbf{B}$.*

Solution. From equation (3.2) we find:

$$m\frac{d\mathbf{v}}{dt} = -e\mathbf{v} \times \mathbf{B}.$$

When the magnetic field has a direction along the z-axis, so that $\mathbf{B} = (0, 0, B)$, and the particle moves only in the transverse plane, with a velocity $\mathbf{v} = (\dot{x}, \dot{y}, 0)$, then the equations of motion become two differential equations:

$$m\ddot{x} = -eB\dot{y}, \quad m\ddot{y} = eB\dot{x}.$$

Their general solution is:

$$x(t) = x_0 - R\sin(\omega t + \phi), \quad y(t) = y_0 + R\cos(\omega t + \phi).$$

A particle moves in a circle, for $B > 0$ counterclockwise, with center at point (x_0, y_0) and radius R, with arbitrary phase ϕ. The circular frequency is constant

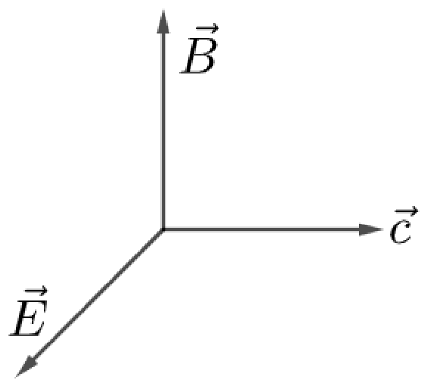
$$\omega = \frac{eB}{m}$$

and is called the cyclotron frequency. □

When the external magnetic field is perpendicular to the velocity of the electron, it is also perpendicular to its acceleration. The acceleration $\dot{\mathbf{v}} = \mathbf{F}/m$ has a

lateral direction to the velocity, in a plane perpendicular to the vector B, so the velocity v of the electron moves along with it. Again, a lateral acceleration occurs, which drags the velocity along with it, and so on in the movement of the electron in a circle around the magnetic field with a frequency proportional to the strength of the magnetic field and the charge and inversely proportional to the mass of the electron.

Solving this example is not worth it for a photon, a particle of light, because it has no charge, no magnetic field, and no mass.



However, photons are electromagnetic radiation, so although we do not consider these elements of them to be real, i.e., they do not exist like, say, a quark, a building block of hadrons (e.g., protons) with spin 1/2 and charge 1/3 e or 2/3 e, we can talk about unreal information in photons. For example, the internal magnetic field of a photon $\vec{B}$ acts on the electric field $\vec{E}$, inducing a lateral motion $\vec{c}$. Fictitious $\vec{B} \times \vec{E} \to \vec{c}$, then $\vec{E} \times \vec{c} \to \vec{B}$ and so on, as in the figure on the left. There is no external permanent magnetic field, but the photon carries it with it, moving at a constant speed $\vec{c}$. This was my digression.

Further features of this concept would be that the internal information of light is not ordered in the aforementioned way but should be viewed as a wave of probability. The aforementioned properties are like fictions, impermanent and, in the absence of choice, forced to constantly emerge and maintain the reality of photons.

Let us now correct the actual formula (3.2) with two more ingredients. First, let us add an electric field, **E**. This will accelerate the charges and, in the absence of a magnetic field, result in a current in the direction of **E**. The second ingredient is linear friction, which should hinder the progress of the electrons, either by impurities, the lattice, or other electrons. We obtain:

$$m\frac{d\mathbf{v}}{dt} = -e\mathbf{E} - e\mathbf{v} \times \mathbf{B} - \frac{m\mathbf{v}}{\tau}. \tag{3.3}$$

The friction coefficient τ is called the scattering time. It can be thought of as the average time between collisions. This is the Drude model (1900) which treats electrons as spheres. In the absence of acceleration, $d\mathbf{v}/dt = 0$, the particle velocity is:

$$\mathbf{v} + \frac{e\tau}{m}\mathbf{v} \times \mathbf{B} = -\frac{e\tau}{m}\mathbf{E}.$$

The current density is proportional to the velocity $\mathbf{J} = -ne\mathbf{v}$, where n is the charge carrier density, so we have:

$$\begin{pmatrix} 1 & \omega\tau \\ -\omega\tau & 1 \end{pmatrix}\mathbf{J} = \frac{e^2 n\tau}{m}\mathbf{E},$$

which is the matrix representation of Ohm's law

$$\mathbf{J} = \sigma \mathbf{E}. \tag{3.4}$$

It describes the flow of current in an electric field. The constant of proportionality σ is the conductivity. This Ohm's law is an improved version in the sense that near a magnetic field σ is not a number, but a matrix. Sometimes we call it the conductivity tensor:

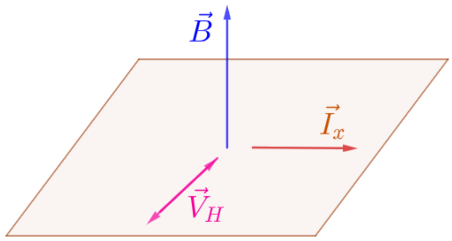
$$\sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ -\sigma_{xy} & \sigma_{yy}. \end{pmatrix} \tag{3.5}$$

From the Drude equation we find an explicit expression for this constant:

$$\sigma = \frac{\sigma_0}{1+\omega^2\tau^2}\begin{pmatrix} 1 & -\omega\tau \\ \omega\tau & 1 \end{pmatrix}, \quad \sigma_0 = \frac{ne^2\tau}{m},$$

with conductivity σ_0 outside the magnetic field. The off-diagonal terms are part of the Hall effect: in equilibrium, a current in the x-direction requires an electric field with a component in the y-direction.

The previous photon image is now on the right, more realistic[1]. A constant external magnetic field $\vec{B}$ affects the motion of the current $\vec{I}_x$ and the Hall voltage $\vec{V}_H$ in a plane perpendicular to it. The applied electric field in the x-direction has a current density $\vec{J}_x$ which is deflected by the magnetic field $\vec{B}$ in the z-direction and bends the Hall voltage vector $\vec{V}_H$ in the y-direction. This continues around the circle in the plane perpendicular to the magnetic field.



From the point of view of the passive[2] appearance of a magnetic field (fiction) as a reaction to a (real) electric field, we note that the "external" vector $\vec{B}$ originates from some electric current $\vec{I}' \perp \vec{B}$ perpendicular to it. It is a bundle of force lines perpendicular to $\vec{I}_x \| \vec{I}'$, which then attracts it in a way that will be explained in the next subheading. Therefore, the vector $\vec{I}_x$ induces a Hall voltage $\vec{V}_H$ and changes direction towards the y-axis. Then it is again attracted to it by a parallel current $\vec{I}''$ such that both are perpendicular to $\vec{B}$ and the cyclic process continues, as in the solution to Example 109. The beam of magnetic force perpendicular to the plane of rotation of the electric current continues to symmetrically curve in all directions, enabling this circular repetition.

Resistivity (symbol ρ), or the electrical resistance of a conductor, measured per unit cross-sectional area per unit length, is one of the important properties of a material. It is a useful quantity for comparing the ability of a material to conduct electricity. High resistance indicates poor conductors. *Resistance* (symbol R), or the ability to be unaffected by something, is generally equal to resistivity.

[1] David Tong: The Quantum Hall Effect; Cambridge, 2016
[2] Unofficial physics, my information theory.

Resistivity is defined as the inverse of conductivity, and this holds when it comes to matrices:

$$\rho = \sigma^{-1} = \begin{pmatrix} \rho_{xx} & \rho_{xy} \\ -\rho_{xy} & \rho_{yy} \end{pmatrix} = \frac{1}{\sigma_0} \begin{pmatrix} 1 & \omega\tau \\ -\omega\tau & 1 \end{pmatrix}. \tag{3.6}$$

This second matrix follows from the Drude model. This resistivity tensor on the side diagonal $\rho_{xy} = \omega\tau/\sigma_0$ is independent of the dissipation time τ and tells us about the independence of the material from the things responsible for the dissipation.

However, we usually measure the resistance R which differs from the resistance ρ by a factor of scale, but in the case of ρ_{xy} the two measurements coincide. For example, let a material of length L be placed along the y-axis and let us apply a voltage V_y in the y-direction to measure the resulting current I_x in the x-direction. The lateral resistance is:

$$R_{xy} = \frac{V_x}{I_x} = \frac{E_y L}{J_x L} = \frac{E_y}{J_x} = -\rho_{xy}.$$

So, the resistance we calculate ρ_{xy} and what we measure, R_{xy}, are equal in this case. But, when we measure the longitudinal resistance R_{xx}, we divide it by the corresponding lengths and extract the resistance ρ_{xx}.

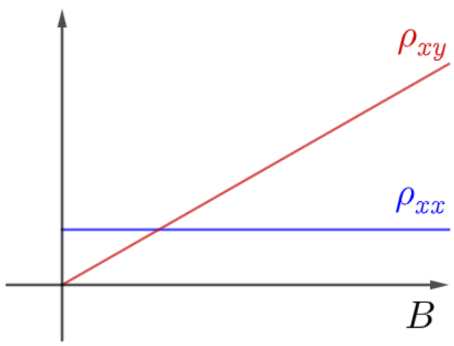
In addition to these two determinants, here is another one. For a current I_x flowing in the x-direction and an associated electric field E_y in the y-direction, the *Hall coefficient* is defined by:

$$R_H = -\frac{E_y}{J_x B} = \frac{\rho_{xy}}{B}, \tag{3.7}$$

which is according to the Drude model:

$$R_H = \frac{\omega}{B\sigma_0} = \frac{1}{ne}.$$

The Hall coefficient depends only on microscopic information about the material, the charge and the density of the conducting particles, and does not depend on the scattering time τ; it is insensitive to all friction processes occurring in the material.



This gives us the predictions for the resistance and the endurance:

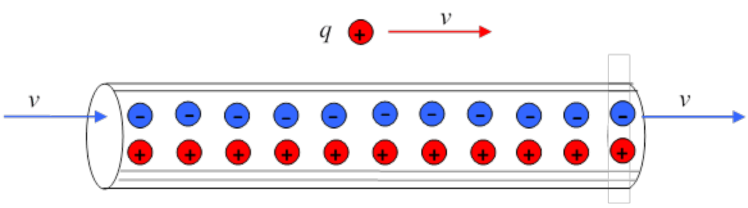
$$\rho_{xy} = \frac{B}{ne}, \quad \rho_{xx} = \frac{m}{ne^2\tau}. \tag{3.8}$$

Notice that only ρ_{xx} depends on the scattering time τ, and $\rho_{xx} \to 0$ as scattering processes become less important and $\tau \to \infty$. If we plot the two resistances as a function of the magnetic field, then the classical (as opposed to the quantum Hall effect) expectation is that they should look like the figure on the left.

3.1.3 Coulomb's law

A moving charge or electric current is a source of a magnetic field. A current of charge or a conductor carrying a current induces a magnetic field around itself, even if it is a single moving electron with a charge of only $e = -1.6 \times 10^{-19}$ C. Since moving charges such as electrons generate a magnetic field, it is fascinating to see what happens to the field for an observer at rest relative to a moving electron from the point of view of the relativity of motion.

This is a seemingly paradoxical question of the relativity of motion (Relativity), to which physics has a successful answer using only the Coulomb force (1785). In short, when we observe a neutral conductor and an electron next to it moving side by side at a speed v, as in the following figure, a magnetic field appears around them that attracts them.



The explanation of this attraction, if we move together with the conductor and the electron, that is, which move but we are at rest between them, is in the rest of the separated electron and the electron in the conductor and in the movement of its protons. The relative contraction of lengths is

$$\Delta\ell = \Delta\ell_0\sqrt{1 - \frac{v^2}{c^2}}, \tag{3.9}$$

where $\Delta\ell_0$ is the length at rest (own length), $\Delta\ell$ is the relative length in the direction of movement, v is the speed of movement, and c is the speed of light in a vacuum. The protons therefore become denser; the electrons do not because we are at rest relative to them, so the opposite charge becomes stronger and the conductor attracts the electron.

The constant central force, already from the consideration of conics (19. task), will decrease with the square of the distance

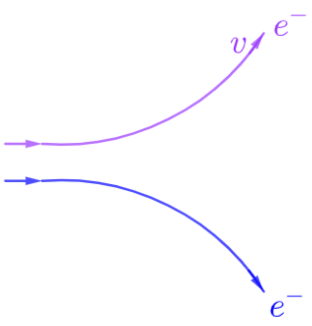
$$F = k_e\frac{q_1 q_2}{r^2},$$

where q_1 and q_2 are the charges, and it is then a matter of measurement to determine the coefficient k_e in a given medium. The invariance of electromagnetic waves using Lorentz transformations is known ([8], Example 1.4.3.), as is the independence of the charge strength from motion.

It turns out that the magnetic force "does not exist," but this is only when the definition of the term "exist" is narrowed down to "reality." However, here we consider both "real" and "unreal" as existing phenomena. The latter are, for example,

our thoughts, which may not be true. Thus magnetism is an existing fictitious, we can even say passive, phenomenon induced by real electricity.

When we observe two lone electrons moving apart, repelled by the Coulomb force, we can always invoke relativity (even though there are no protons) and say that they are moving by inducing a magnetic field that is also attracting them. They will therefore not move apart fast enough, as the Coulomb force would predict. But this "relativistic paradoxčan also be explained in the previous way – as fictitious magnetism.



Let's imagine that we are at rest with respect to one of the electrons, e.g. the lower one in the picture on the left, and we observe the upper one moving away with a velocity v. At that moment, the relativistic time dilation is

$$\Delta t = \frac{\Delta t_0}{\sqrt{1 - \frac{v^2}{c^2}}}, \tag{3.10}$$

where Δt_0 is proper time, and Δt is relative time. We therefore see its motion slowed down, and we will attribute the decrease in the Coulomb force to the magnetic field.

I wrote about it in more detail in my blog (Induction), but answering the question posed and perhaps not emphasizing the unreality of magnetization enough. There you can also find (an attempt) to interpret the force between electrons using principled minimalism (Current).

From the example where electricity produces magnetism and the first is “reality” and the second is “fiction,” we glimpse something of the enormous layering of reality, or the similar and yet quite different complexity of its secondary phenomena. When we use a compass to see the magnetism of the Earth, we measure a secondary thing through technical intermediaries. We read “unreal” forms by relying on “unreal.” All around us are such energies, whose only total quantities are conserved and whose forms change (kinetic, potential, chemical), so when we get burned in a fire, we still feel its temporality very much.

These examples also tell us something about synergy, about the emergence and possibility of inanimate substances to bind with fictions into living beings that then have so much freedom that we can even lie. In the picture on the right is a picture of collective ants with a story about (Ant Colony) their small individual powers and extremely simple brains the size of grains of sand, which together can work wonders. They are an example of giving up personal information (the amount of options) for the sake of collective information, which thus gains vitality and in return offers security to its members.

I wrote earlier about the tendency to get rid of excess options and the principle of minimalism. Hence the more frequent realization of certain outcomes, their favoring due to the force of probability, which is attractive, its opposite the repul-

sive force of uncertainty. That is why, throughout evolution, we have been driven by the fear of the unknown and the need for security, then the desire for laziness or efficiency, and the search for order from social to natural laws.

We can now go a step further in defining the "quantity of fictions" along the lines of energy and its forms, implying only the secondary effects generated by the primary. The total quantity of such phenomena, primary with all secondary, is constant. We say it is equal to the quantity of the basic phenomenon, the whole reality that generates them. In particular, the entire quantity of the electromagnetic field in a given closed system comes from the energy of electricity. Classically, it is the form of energy produced by the flow of charged particles, such as electrons, through a conductor.

For example, writing about a lie as an unreal phenomenon (1.3 Truth), to the knowledge that the "world of inaccuracy" is equivalent to the "world of accuracy," we can add that the quantity of the latter is twice that same quantity, which means that we are dealing with infinities. However, the total energy contained in, say, a liter of water is not infinite in physics problems in which we work with it, but it is far greater than what we usually get in solutions.

The breadth of research options in that relatively small amount of this simple substance (it is easy to imagine some others besides water) is also understood from chemistry itself, to which the following two examples are dedicated. The next two are there to see even a hint of the possibilities of physics, and we will see how infinite energies lie behind a single electron in expression (3.13). Their purpose is to bring us closer to understanding infinity as a type of reality, just like frequencies or colors of light that we cannot see or sounds that we cannot hear.

So, the first example is the specific heat of water of 4182 Joules/kg. Its density is about 1 kg/liter, so it takes 4.128 kJ (of energy) to heat 1 liter of water by 1 degree Celsius, or vice versa, as much as we can get by cooling it. In determining this, we have successfully ignored, for example, the enormous energy of the atoms lurking in the structure of water, which would actually far overshadow the magnitudes in this problem. The example is instructive and simple, and the next one is there to give us some sense of magnitude for the third one concerning the hydrogen atom and to prepare us for a slightly more complex insight.

Another example is the electronvolt. It is a unit of energy equal to the kinetic energy acquired by a free electron in a vacuum when passing through a potential difference of one volt. One electronvolt is a very small unit of energy and is approximately $1eV = 1.602\times10^{-19}$ J. In the Bohr model of the atom, on the n-th orbit, after calculating and substituting constants, we find for the hydrogen atom the electron energy $E_n = -(13.6\cdot\text{ eV})\frac{1}{n^2}$, $n = 1, 2, 3, \ldots$.

Example 110. *A hydrogen atom from the ground state ($n = 1$) absorbs a photon and transitions to the state $n = 3$. What is the energy of the photon?*

Solution. The ground state energy of a hydrogen atom is $E_1 = -(13.6\cdot\text{ eV})/1^2 = -13.6$ eV, and the third state energy is $E_3 = -13.6/3^2 = -1.51$ eV. The energy of the absorbed photon is the energy level change $\Delta E_{13} = E_3 - E_1 = -1.51 - (-13.6) = 12.09$

eV. If we include $1eV = 1.602 \times 10^{-19}$ joules, the energy is $\Delta E_{13} = 12.09 \cdot 1.602 \times 10^{-19}$ J, or $\Delta E_{13} = 19.368 \times 10^{-19}$ J. □

Knowing the photon energy, rounded to $\Delta E_{13} = 1.94 \times 10^{-18}$ J, from the equation $\Delta E = hf$, where $h = 6.626 \times 10^{-34}$ J is Planck's constant, we find the photon frequency $f = 2.93 \times 10^{15}$ Hz, and then from $\lambda f = c = 300 \times 10^6$ m/s we also find the wavelength of the absorbed photon $\lambda = 1.02 \times 10^{-7}$ m.

We have already dealt with the wavelengths from the example here (2.2.4 Spectrum) in a slightly different way. For the photon emission spectrum, for the jump of an electron from the state $n_1 \to n_2$ in the hydrogen atom, the formula is given:

$$\frac{1}{\lambda} = R\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right), \tag{3.11}$$

where $R = 10\ 973\ 731.6$ m^{-1} is the Rydberg constant, and λ is the photon wavelength. Checking the result from the previous example, setting $n_1 = 1$ and $n_2 = 3$, we calculate $\lambda = 1.02 \times 10^{-7}$ m. This is the same result (to three significant figures) as the previous one. I have calculated these energy formulas in detail earlier [10] (1.4.13 The Hydrogen Atom), solving the Schrödinger equation.

Example 111. *Let's find the wavelength of the photon that will be absorbed if the electron of a hydrogen atom is raised from the state $n = 3$ to the state $n = 5$.*

Solution. The energies of electrons in the 3rd and 5th states (orbits) are:

$$E_3 = -(13.6 \cdot eV)\frac{1}{n^2} = -13.6 \cdot 1.6 \times 10^{-19} \cdot \frac{1}{3^2} \text{ J} = -2.42 \times 10^{-19} \text{ J},$$

$$E_5 = -(13.6 \cdot eV)\frac{1}{5^2} = -13.6 \cdot 1.6 \times 10^{-19} \cdot \frac{1}{25} \text{ J} = -0.87 \times 10^{-19} \text{ J}.$$

The energy difference is $\Delta E_{35} = E_5 - E_3 = 1.55 \times 10^{-19}$ J, so for the frequency and wavelength ($f\lambda = c$) of the absorbed photon we now find:

$$f = \frac{\Delta E_{35}}{h} = \frac{1.55 \times 10^{-19} \text{ J}}{6.626 \times 10^{-34} \text{ J s}} = 2.34 \times 10^{14} \text{ Hz},$$

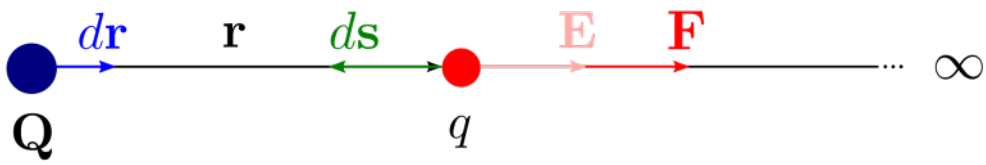
$$\lambda = \frac{c}{f} = \frac{300 \times 10^6 \text{ m/s}}{2.34 \times 10^{14} \text{ Hz}} = 1.28 \times 10^{-6} \text{ m}.$$

This is infrared light. □

Checking the result using equation (3.11) with the Rydberg constant, setting $n_1 = 3$ and $n_2 = 5$, gives $\lambda = 1.28 \times 10^{-6}$ m. This is again the same result as calculated in the example. Note that the formula speaks of a decreasing photon energy required to raise an electron from a lower orbit to a higher orbit with the same number of steps. For an electron to jump from the first to the third orbital of a hydrogen atom, the photon energy $E_{13} = 19.368 \times 10^{-19}$ J (joules) needs to be added to it, and for a jump from the third to the fifth, only $\Delta E_{35} = 1.55 \times 10^{-19}$ J.

3.1.4 Potential energy

Electric potential energy is also expressed in joules and is the result of conservative Coulomb forces. It is calculated for a given set of point charges within a given system. An object is said to have electric potential energy based on its own electric charge or from its relative position with respect to other electrically charged objects.



The figure shows the total charge Q of some particles, from which a given charge is at a distance of $\mathbf{r}$. According to Coulomb's law, the electric potential $V(\mathbf{r})$ between them and the electrostatic potential energy of the charge q with respect to Q are:

$$V(\mathbf{r}) = k_e\frac{q}{r}, \quad U_E = \frac{1}{4\pi\epsilon_0}\frac{qQ}{r}, \tag{3.12}$$

where $r = |\mathbf{r}|$ is the distance between the charges. These can be added together, and, for example, the electrostatic potential energy in a system of three charges Q_1, Q_2 and Q_3 is:

$$U_E = \frac{1}{4\pi\epsilon_0}\left(\frac{Q_1Q_2}{r_{12}} + \frac{Q_1Q_3}{r_{13}} + \frac{Q_2Q_3}{r_{23}}\right)$$

where r_{ij} is the distance from the ith to the jth charge.

Example 112. *The energy within the electrostatic field arrangement* $\mathbf{E}$ *in a vacuum of permittivity* ϵ_0 *has the density:*

$$u_e = \frac{dU}{dV} = \frac{1}{2}\epsilon_0|\mathbf{E}|^2.$$

This V is now the volume.

Proof. We get this from Gauss's law, from Maxwell's 1st equation:

$$\nabla\cdot\mathbf{E} = \frac{\rho}{\epsilon_0},$$

or integrally using the Φ electric potential (the amount of work, energy required per unit of electric charge to move the charge from a reference point to a certain point in an electric field):

$$U = \frac{1}{2}\int_S \rho\Phi(r)\,dV = \frac{1}{2}\int_S \epsilon_0(\nabla\cdot\mathbf{E})\Phi\,dV,$$

where the integrals are over the entire space S. Using the divergence property:

$$\nabla\cdot(\vec{v}\alpha) = (\nabla\cdot\vec{v})\alpha + \vec{v}\cdot(\nabla\alpha) \;\Rightarrow\; (\nabla\cdot\vec{v})\alpha = \nabla\cdot(\vec{v}\alpha) - \vec{v}\cdot(\nabla\alpha),$$

we find:

$$U = \frac{\epsilon_0}{2}\int_S \nabla\cdot(\mathbf{E}\Phi)\,dV - \frac{\epsilon_0}{2}\int_S (\nabla\Phi)\cdot\mathbf{E}\,dV.$$

Using divergence theorem (Gauss's, as well as Ostrogradsky's theorem), which relates the flux of a vector field through a closed surface to the divergence of the field in a closed volume, and taking the area at infinity $\Phi(\infty) = 0$ and putting $\nabla\Phi = -\mathbf{E}$:

$$U = \frac{\epsilon_0}{2}\int_{S'} \Phi\mathbf{E}\cdot d\mathbf{A} - \frac{\epsilon}{2}\int_S (-\mathbf{E})\cdot\mathbf{E}\,dV = \int_S \frac{1}{2}\epsilon_0|\mathbf{E}|^2\,dV,$$

where S' are the boundaries of the space, $\mathbf{A}$ is the area vector (Eng. Area) and the first integral is zero, and the second, where S is the entire space, gives the desired result. □

The square of the electric field intensity is the number $|\mathbf{E}|^2$, the result of the scalar multiplication of the field vector with itself, U is the energy of the electric field, and dV is the infinitesimal volume element.

Let us further assume that we have two charges Q_1 and Q_2. Each charge has its own electric field $\mathbf{E}_1$ and $\mathbf{E}_2$, the total electric field being the vector sum $\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$. The total potential energy "stored in the fields" in this configuration can be divided into three parts:

$$U = \frac{\epsilon_0}{2}\int_S |\mathbf{E}_1 + \mathbf{E}_2|^2\,dV = \frac{\epsilon_0}{2}\int_S (|\mathbf{E}_1|^2 + |\mathbf{E}_2|^2 + 2\mathbf{E}_1\cdot\mathbf{E}_2)\,dV = U_1 + U_2 + U_{12}.$$

When we integrate over the entire space S, with the charges separated by r, the third of the sums of the results of the above integrals is:

$$U_{12} = \frac{Q_1Q_2}{4\pi r},$$

which is the sum of the second of the formulas (3.11). However, the integrals U_1 and U_2 diverge. Namely:

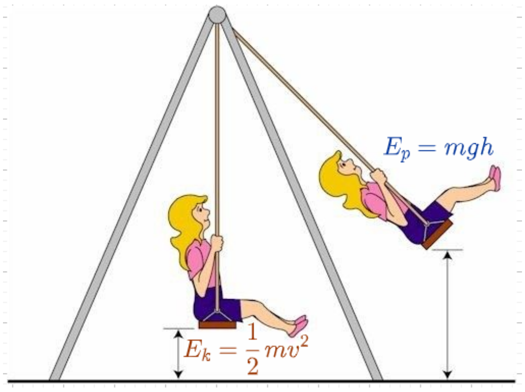
$$U_1 = \frac{\epsilon_0}{2}\int_S \left(\frac{Q_1}{4\pi r^2}\right)^2 r^2\sin\theta\, dr\,d\theta\,d\phi = \frac{Q_1^2}{8\pi\epsilon_0}\int_0^\infty \frac{1}{r^2}\,dr = \frac{Q_1^2}{8\pi\epsilon_0}\frac{1}{r}\bigg|_0^\infty, \tag{3.13}$$

and this grows indefinitely due to the lower bound. Similarly, U_2 diverges.

We usually skip over this divergence because we don't care about the absolute value of the potential energy; only the differences between the potential energies of the configurations matter. The quantities U_1 and U_2 do not depend on the location of the other charge, so they are amounts of energy that each charge somehow carries with it, but we reset our žero"to the potential energy for the separation of the system, $r \to \infty$, and then our new potential energy is U_{12} by itself, and everything is fine.

Four common examples of potential energy include a skydiver waiting to jump out of an airplane, a rubber band stretched between two fingers, water sitting behind a dam, and the energy in a battery. All four examples are examples of stored

energy that has the potential to do work upon its release (Study.com). On the other hand, potential energy is another convenient "information coupling"topic for explaining the differences between *real and fictitious*.



Potential energy, in the picture on the left of the person on the swing $E_p = mgh$, is a fiction about something that does not exist, but we certainly know what it will degenerate into if the right circumstances are met. Here, in the case of gravity, moving from a higher to a lower swing position, potential energy develops into kinetic energy ($E_k = \frac{1}{2}mv^2$).

The sum of kinetic and potential energy is *mechanical energy*. The mechanical energy of an object may be more the result of its motion (kinetic energy) or the manifestation of its stored energy of position (potential energy), but their sum is a constant result, and that is their reality.

Mechanical energy is such a real phenomenon that even in microphysics we do not find a violation of its duration (2.2.1 Hamiltonian). This applicability tells us about the obligation to consider its very universality, or similar repeatable schemes of logic of the first order, as a kind of reality. This is how we again arrive at the understanding that a phenomenon is "real" when it is: (1) perceptible, (2) sustainable, and (3) true. Then, based on the Skolem's paradox, we conclude that (some) infinities are also realities for us.

Skolem's paradox arises from a part of the Löwenheim–Skolem theorem; Skolem was the first to notice the absurd claim of that theorem that every model of logic, if consistent, has an equivalent model that is countable. This seems contradictory, since Cantor had proved that there are sets that are not countable. So the apparent contradiction is that a model that is countable in itself, and therefore contains only countable sets, satisfies the first-order sentence that intuitively says, "There are uncountable sets."We add[3] analogous to finite reality telling us "there is an infinite reality."

Narrowing these analogies down to the smaller environment of our local perceptions, in a real-world scene of the swinging of the person in the picture above, or such a load, the swing would spontaneously and slowly stop. In the aggregate mechanical energy, there is a loss of energy through friction, but the total energy of a larger scope, say a *closed system*, is then conserved. This escape through friction is just one way that nature tells us "there is a reality beyond the local."

3.1.5 Linking expressions

Maxwell's equations can also be written using electromagnetic potentials with a scalar φ and a vector potential $\mathbf{A}$:

[3]My addition, not in official science.

1. $\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \iff \nabla \cdot (-\nabla\varphi - \frac{\partial \mathbf{A}}{\partial t}) = \frac{\rho}{\epsilon_0}$;
2. $\nabla \cdot \mathbf{B} = 0 \iff \nabla \cdot (\nabla \times \mathbf{A}) = 0$;
3. $\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \iff \nabla \times (-\nabla\varphi - \frac{\partial \mathbf{A}}{\partial t}) = -\frac{\partial \mathbf{B}}{\partial t}$;
4. $\nabla \times \mathbf{B} = \mu_0 \cdot \mathbf{J} + \mu_0 \cdot \epsilon_0 \cdot \frac{\partial \mathbf{E}}{\partial t} \iff \nabla \times (\nabla \times \mathbf{A}) = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial}{\partial t}(-\nabla\varphi - \frac{\partial \mathbf{A}}{\partial t})$.

The electric field expressed in terms of the scalar potential φ, and the magnetic field expressed in terms of the vector potential are, respectively:

$$\mathbf{E} = -\nabla\varphi - \frac{\partial \mathbf{A}}{\partial t}, \quad \mathbf{B} = \nabla \times \mathbf{A}. \tag{3.14}$$

The first of Maxwell's four equations is Gauss's law. Here is how it comes from Coulomb's law, which follows from the consideration of conics (19. task), that the constant central force decreases with the square of the distance and is proportional to the charges, and the scaling factor is a matter of measurement.

Example 113. *Let's derive Gauss's law from Coulomb's.*

Solution. The Coulomb force $\mathbf{F}_1$ from the first point charge of q_1 on the second q_2 is directly proportional to each of the two charges and inversely proportional to the square of the distance $r = |\mathbf{r}_2 - \mathbf{r}_1|$ between them. These are vector quantities:

$$\mathbf{F}_1 = k_e \frac{q_1 q_2}{r^2} \mathbf{r}_{12}, \quad \mathbf{r}_{12} = \frac{\mathbf{r}_1 - \mathbf{r}_2}{|\mathbf{r}_1 - \mathbf{r}_2|}.$$

Here $\mathbf{r}_{12}$ is a unit vector directed from q_2 to q_1, and k_e is Coulomb's constant of proportionality. The electric field at location q_1 due to q_2 is

$$\mathbf{E} = k_e \frac{q_2}{r^2} \mathbf{r}_{12}.$$

We assume that an electric field can be a collection of charges by the vector sum of the electric fields produced by each of the charges in the collection. This is known as the superposition principle:

$$\mathbf{E}(\mathbf{r}) = \sum_{i=1}^{n} \mathbf{E}_i, \quad \mathbf{E}_i = k_e \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|^2} \mathbf{r}_{0i}, \quad \mathbf{r}_{0i} = \frac{\mathbf{r} - \mathbf{r}_i}{|\mathbf{r} - \mathbf{r}_i|}.$$

If the charge in the system can be considered to be continuously distributed over a volume V, the sum becomes an integral:

$$d\mathbf{E} = k_e \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} (\mathbf{r} - \mathbf{r}')\, dV, \quad \mathbf{E}(\mathbf{r}) = \int_V k_e \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} (\mathbf{r} - \mathbf{r}')\, dV.$$

Gauss's law provides the basis for determining the electric flux through a closed surface. This is done for a single charge and the result is extended to a system of charges through the superposition principle, with a point charge q at the origin

of the coordinate system. The following are the formulas for the electric flux, Φ, through a surface S, due to a charge at a point:

$$d\Phi = \mathbf{E}\cdot\mathbf{n}\,dS, \quad \oint_S d\Phi = \oint_S k_e \frac{q\mathbf{r}\cdot\mathbf{n}}{r^2}\,dS,$$

where dS is an infinitesimal element of the surface surrounding the charges, and $\mathbf{n}$ is the unit vector normal to that surface.

Next, we use the solid angle theorem Ω. This is the steradian which is defined as "the solid angle at the center of a sphere of radius r subtended by the portion of the surface of a sphere of area r^2." Since the surface area of this sphere is $4\pi r^2$, the definition implies that the sphere has 4π steradians. It is here composed of a closed surface S from a point P, so $\Omega = 4\pi$ when P is inside S, and $\Omega = 0$ if P is outside S. Hence, the charge outside the surface, if any, does not produce an electric current through the surface. It follows from the superposition principle that the flux of the total electric charge Q through the closed surface is $\Phi = 4\pi k_e Q$. This is Gauss's law:

$$\oint_S \mathbf{E}\cdot\mathbf{n}\,dS = 4\pi k_e Q.$$

At a point inside a continuous charge distribution $\rho \neq 0$, a small closed surface S can be chosen that encloses the point, such that $\rho = 0$ throughout the interior of S. From Gauss's law for the surface S and the divergence theorem, we obtain:

$$\oint_S \mathbf{E}\cdot\mathbf{n}\,dS = \int_V \nabla\cdot\mathbf{E}\,dV = 4\pi k_e \int_V \rho\,dV,$$

where V is the volume enclosed by the surface S. Hence

$$\int_V (\nabla\cdot\mathbf{E} - 4\pi k_e\rho)\,dV = 0$$

on an arbitrarily small neighborhood of any point inside the charge distribution. The continuity of the integrand gives

$$\nabla\cdot\mathbf{E} = 4\pi k_e\rho.$$

In SI units,

$$k_e = \frac{1}{4\pi\epsilon_0},$$

where ϵ_0 is the permittivity of vacuum. Thus, Maxwell's first equation

$$\nabla\cdot\mathbf{E} = \rho/\epsilon_0$$

is called Gauss's law. □

These are long-known things[4], and if we were to add something, it could be the momentum operator $\hat{p} = -i\hbar\nabla$. Gauss's law then states that the "scalar product"

[4] I learned it somewhere from Dale Gray, a former professor of physics (1996–2008).

of the momentum process and the electric field is proportional only to the field density. If we remember that we interpret this product as information about perception, then we are talking about the possibility of communication (interaction) with the amount ρ/ϵ_0. That is the first Maxwell equation, and the second would tell us that these impulses do not communicate with the magnetic field. The magnetic field is a secondary phenomenon of the electric, and these momentum processes simply do not notice them. Not everyone will communicate with everyone.

As for the third and fourth equations, recall that we interpret $\hat{p}\times\mathbf{E}$ as a commutator or "surface" similar to Kepler's second law. Unlike the inner product ($\hat{p}\cdot\mathbf{E}$), this outer product speaks of external influences on the coupled ones. Indeed, (equation 3) the outer product of momentum and electric field is proportional to the change in magnetic field, and (equation 4) the outer product of momentum and magnetic field is proportional to the change in electric and electric currents.

Let's try something even newer. In this theory of information, everything is about a kind of give and take. We exchange through interactions what we could not otherwise, hoping for greater certainty. In this way, the electron communicates with measuring devices, handing over the possibilities of the "now" so that its path "then" becomes more certain. By measuring, or observing, the electron is discharged of its uncertainty, with which it regains its previous path – the founders of quantum mechanics (Heisenberg, 1925) noticed with amazement, and we can only now say – because by measuring it hands over its uncertainty to the equipment, after which its path remains more certain.

This reveals the possibility that the present affects the past and, therefore, that the past can change at least slightly. This minimal change, however, is necessary in the idea of "give and take" for the past to be able to send any messages to the future. By allowing us to remember it (archaeologically, forensically, and in other ways), it leaves us information about itself, supplementing the present with the amount it takes away from it by stretching it. This makes it appear paler to us the further it is from us, building its "illusion" of perfection.

The same story of "giving and taking"is easily transferred to abstract truths and laws of physics, with even less tolerance. Information travels in quanta of action $\Delta E\cdot\Delta t$. Thus, packages are stored, which have a variable share of energy ΔE and time Δt. The laws of time endure (and are therefore "real"), but in such a way that time stands still for them ($\Delta t\to\infty$ and therefore $\Delta E\to 0$), making themselves very attractive to everyone with whom they can communicate, who recognizes them. However, the slower time is, the more abstract, less knowable and elusive, and less vital these truths tend to be. Humans also communicate on a cognitive level and are recognized that way, but animals are predominantly emotional and instinctive, and inanimate physical creatures unconditionally follow only the principle of least action, blindly adhering to the laws they "recognize."

3.2 Momentum

Momentum is otherwise defined in physics as the amount of motion multipli-

ed by the amount of mass displaced and the speed at which the mass is traveling. When you walk, run, or do anything else, you create momentum. If both a bicycle and a car are traveling at the same speed down a street, the car will have more momentum. The SI unit for momentum is kg m/s, and the quantity is a vector quantity $\vec{p} = m\vec{v}$.

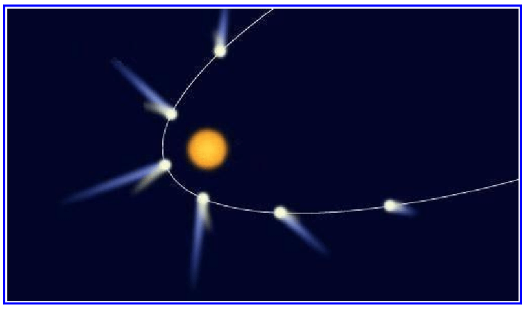
Impulse is a term that refers to the product of the average net force acting on an object over a period of time. When separated from momentum, impulse is denoted by the letter J and is measured in Newton seconds or kilograms per second, $\vec{J} = \vec{F} \cdot \Delta t$. We use the term "momentum"to measure an object's resistance to stopping, and "impulse"refers to the quantity that expresses the effect of the net force applied to the object. Otherwise, the terms are interchangeable.

For example, despite its small mass, a bullet has significant momentum due to its extremely high velocity. However, when we kick a ball, it has a rapid change in momentum and develops a large impulsive force.

3.2.1 Light

How is it possible that light has no mass but has momentum – a question that was one of the first to be addressed by Richard Feynman (1918 - 1988), the founder of quantum electrodynamics and Nobel Prize-winning physicist.

Popular descriptions of his interpretation are increasingly better (light momentum), and are based mainly on a few facts. The first is the observation of comet tails, as shown in the image to the right. The comet tail and coma are visible features of a comet when illuminated by the Sun and can become visible from Earth when the comet passes through the inner Solar System. As the comet approaches the Sun, solar radiation ejects volatile materials from the comet, carrying dust with it.



Most comets are too faint to be seen without the aid of a telescope, but a few each decade are bright enough to be visible to the naked eye, with dust blown straight away by sunlight. Perhaps less obvious to the layman but equally compelling is laboratory evidence of light momentum, the pressure that photons exert on the objects they strike. And momentum is precisely “the ability to push and move,” Feynman said.

The second proof comes from Minkowski's 4-dim space-time model, which was developed for the special theory of relativity. The square of the interval between infinitesimal events (at given times at given places) is

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2. \tag{3.15}$$

This is a pseudo-Euclidean form of the Pythagorean theorem with consequences for other "geometric"properties of relativistic physics that have, of course, been

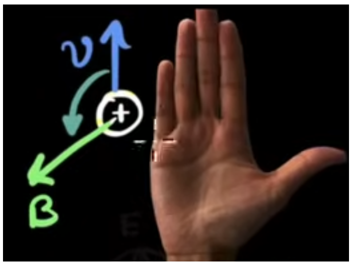
shown to be true in practice. First of all, it is the 4-momentum of Einstein's theory, where there is an additional, fourth component of momentum E/c, and the other three are from classical mechanics, in rectangular coordinates p_x, p_y and p_z. Then we are left with $p_x^2 + p_y^2 + p_z^2 = (mc)^2$ with

$$p^2 = \frac{E^2}{c^2} - p_x^2 - p_y^2 - p_z^2,$$

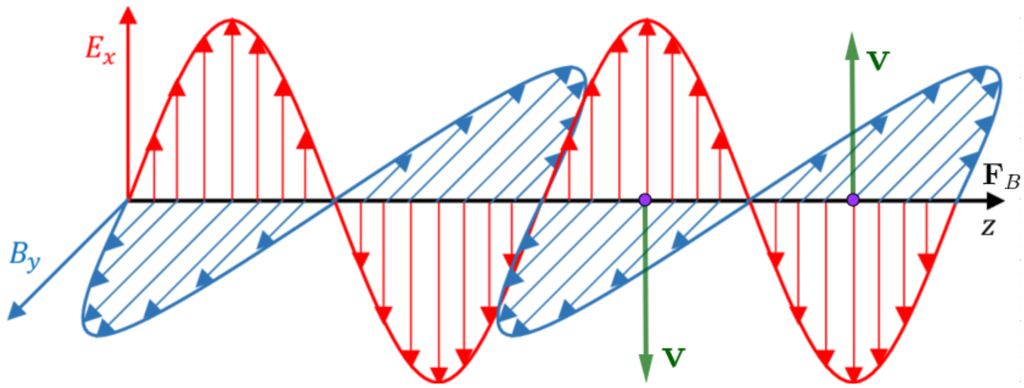
and hence

$$E^2 = m^2c^4 + p^2c^2. \tag{3.16}$$

When $m = 0$, then $E = pc$, and here p is the relativistic momentum of massless light. So, theoretically everything is also fine, so let's see how this will affect the electromagnetic waves of Maxwell's equations, which we mentioned are invariant to Lorentz transformations, otherwise the basis of the special theory of relativity.



The figure on the right shows the *right-hand rule* in the case of the Lorentz force $\mathbf{F}_B = q(\mathbf{v} \times \mathbf{B})$. The fingers are in the direction of the velocity $\mathbf{v}$, the palm is facing the magnetic field $\mathbf{B}$, and the thumb is facing the direction of the force $\mathbf{F}_B$. In this case, the negative charge q reverses the direction of the force vector, so when we apply this vector multiplication of vectors to the electromagnetic wave (light), we get a force in the direction of light motion, a force that arises from the translational (perpendicular) motion of the electric field, the real phase of the wave that induces the fictitious magnetic phase, and acts on the longitudinal (longitudinal) motion of the photon.



The electric force $\mathbf{F}_E = q\mathbf{E}$ on a negative charge q acts in the opposite direction to the electric field $\mathbf{E}$. When the phase of the electric field is down, as in the previous figure, then the velocity $\mathbf{v}$ is going up, the magnetic field is there (opposite the ordinate direction), and the magnetic force $\mathbf{F}_B$ is to the right. The previous velocity, in the figure, is the downward motion of the upper electric field that induces the magnetic field here (in the direction of the y-axis), and the magnetic force $\mathbf{F}_B$ again to the right.

This is Feynman's explanation of the longitudinal motion of vertically oscillating light. It is accompanied by an interpretation using Maxwell's equations, which is combined with relativity and quantum mechanics, unexpectedly for the great theories of physics that emerged later.

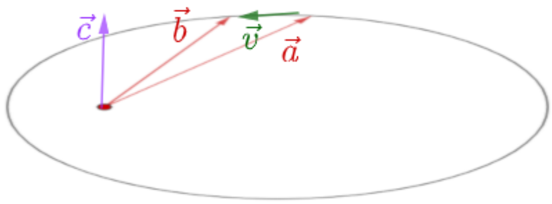
In addition to the 4-dim treatment of space-time (3.15), these two new fields of physics brought us the interpretation of vectors as physical states, and of processes as linear operators. Hence, momentum and energy processes look similar to momentum and energy, with quantum operators as follows:

$$\hat{p} = -i\hbar\nabla = -i\hbar(\mathbf{e}_x\frac{\partial}{\partial x} + \mathbf{e}_y\frac{\partial}{\partial y} + \mathbf{e}_z\frac{\partial}{\partial z}), \quad \hat{E} = i\hbar\frac{\partial}{\partial t}. \tag{3.17}$$

We recognize energy as the fourth coordinate of momentum, $\hat{p}_t = -\hat{E}/c$, perpendicular to the other three axes. In this way, the first Maxwell equation takes on the meaning of the inner (scalar) product of momentum and electric field. Since this is proportional to the electric field density, ρ/ϵ_0, the electric field is not perpendicular to the momentum but is slightly tilted in the direction of the momentum. This is a novelty, necessary to interpret the motion of light waves consistently with the previous one and using Maxwell's equations.

The second Maxwell equation tells us that the momentum process of a photon and the magnetic field induced in it are mutually perpendicular, $\hat{p}\cdot\mathbf{B} = 0$. The third, that the external, vector product of the momentum vector and the electric field ($\hat{p}\times\mathbf{E}$) depends on changes in the magnetic field ($\partial_t\mathbf{B}$), and the fourth that the external product of the momentum and the magnetic field changes with the sum of the current and the change in the electric field. The momentum is the process that again drives the light wave longitudinally (longitudinal) as it oscillates translationally (transversely).

This is a brief explanation of *cross product* vectors for the case of a field $\vec{c} = \vec{a}\times\vec{b}$ when a charge is moving at a velocity $\vec{v}$, as in the figure on the right. Thus, the electric field induces a magnetic field, and then vice versa, which in this way creates a force and momentum for movement at a velocity $\vec{v}$. The velocity changes, but the area swept by the vector $\vec{a}\to\vec{b}$ is always equal:



$$|\vec{c}| = [\vec{a},\vec{b}]_z = a_x b_y - a_y b_x = |\vec{a}||\vec{b}|\sin\angle(\vec{a},\vec{b})$$

in equal times. The trajectory described by the charge is a conic, in the case of a constant central force. These conics are ellipses and parabolas when the forces are attractive, or hyperbolas when the forces are repulsive, or straight lines when the forces are zero, and with all this the intensity of such a force (constant central) decreases with the square of the distance.

I have explained and proven this in detail starting from the chapter "1.2.4 Conic Sections"in the ten or so pages of the appendix there and, I hope, in sufficient detail

that there is no need to add anything to it now. However, we will have a lot more to say about the commutator $[\vec{a},\vec{b}]_z$ as the angular momentum.

Before that, let us recall the discussion in section "2.4.2 Uniqueness", that *subjects' perceptions* will always be different, even if they are the same forms considered here. The same is true of light. It enters the environments of different observers in different directions, depending on the distance or approach of the subject to its source, it has a lower or higher frequency, a longer or shorter wavelength, but its speed in vacuum always remains the same.

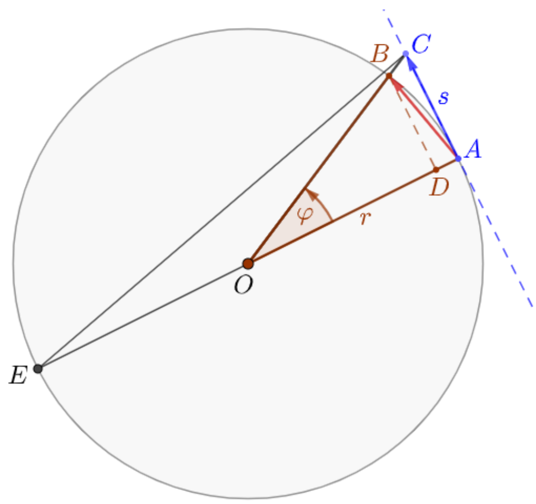
However, when the velocity of an object thrown does not depend on the velocity of the vehicle from which it was thrown, then time stands still for that object. The subject of observation moves in space-time such that his present progresses toward the future at some constant (inherent) velocity, which is actually the speed of light c. That is the observation we need on this occasion.

It is known, on the other hand, from the calculation of time dilation (3.10) of a system moving with velocity v relative to the observer. When $v \to c$, the time of the observed object slows down more and more and in the limiting case it stops.

So, what subjects see as light are different positions from different moments from their own time stream but within a common part of the space-time continuum, otherwise never exactly equally perceived. The very existence in all those times speaks of the reality of the photon, and the continuum of possible outcomes in the countable infinity of a single time stream is a confirmation of additional dimension of time.

3.2.2 Spinner

Rotational momentum (or moment of momentum, or angular momentum) is a physical quantity that measures the tendency of a material body to continue rotating. What lasts would last, a paraphrased consequence of the principle of minimalism and, on the other hand, the conservation of angular momentum.



The image on the left shows the center O of a circle of radius r, which is touched by a dashed line at point A. On the line, tangent to the circle, is the length $s = AC$, which is seen from O at an angle of φ. On the line OC intersecting the circle is the point B, from which the normal to OA is drawn to point D. The line AO intersects the same circle at point E, making AE the diameter of the circle of length $2r$.

The circle rotates uniformly with an angular velocity $\omega = \text{const}$, so $\varphi = \omega t$ after time t. Point A will arrive at B of the circle after that time by traveling the

path $AB = vt$ at some constant velocity v, overcoming the acceleration a with which it can travel the path $AD = \frac{1}{2}at^2$, because it does not move by inertia along the tangent $A \to C$. For very small times $t \to 0$, the arc and the length AB can be considered equal, so from what has been established we have $AB : AD = AE : AB$, i.e. $AB^2 = AD \cdot AE$, and hence:

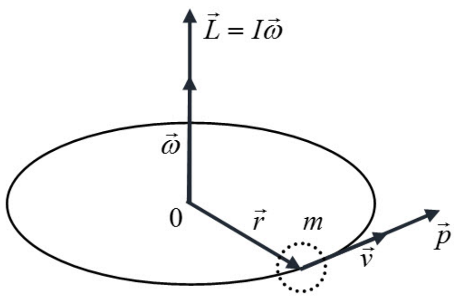
$$a = \frac{v^2}{r} = r\omega^2, \quad F = mr\omega^2, \tag{3.18}$$

where F is the *centrifugal force* that pulls the material point A of mass m from O.

There is a permanent physical connection that does not allow the point A to leave the circle and it does not allow the top, whirligig or gyroscope, to stop rotating around its axis when rotated. The area of the circular section $\Pi_1(OAB) = \frac{1}{2}r^2\varphi$ is less than the area of the triangle $\Pi_2(OAC) = \frac{1}{2}r^2 \operatorname{tg}\varphi$ by the same amount $\Delta\Pi = \Pi_2 - \Pi_1 = \frac{1}{2}r^2(\operatorname{tg}\varphi - \varphi)$ during equal time intervals t. Wherever the segment s is of the tangent, it is the base of a triangle with height OA and area Π_2, thus sliding along the tangent represents inertial motion in the absence of a force from point O.

However, there is a force connecting the point A to the circle that creates a constant area deficit $\Delta\Pi$ that causes that point to rotate, which we call the angular momentum

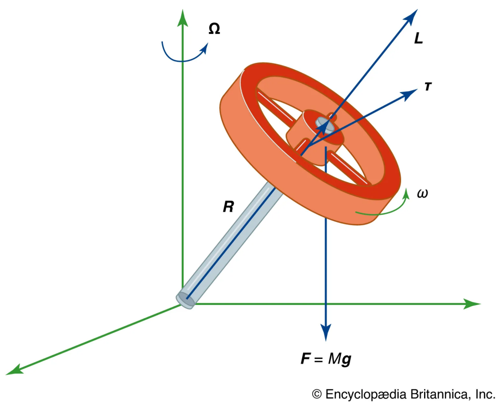
$$\vec{L} = \vec{r} \times \vec{p}, \tag{3.19}$$



where $\vec{r}$ is the position vector of the particle A, relative to the origin O, and $\vec{p}$ is the momentum of the particle. The angular momentum vector is equal to the external, or vector product, of the position vector and the momentum of the particle. The figure on the right also shows

$$\vec{L} = I\vec{\omega},$$

where I is the moment of inertia of the body (in the general case a tensor quantity), and $\vec{\omega} = \vec{\varphi}/t$ is the angular velocity vector now written using the oriented angle.



These šurface deficiency"intrusions are additions to my information theory in the otherwise familiar realm of classical physics. They are easily recognized further in more complex rotations about a moving axis (tumbling), for example, an axis rotating about another axis as in the figure on the left.

Slightly more complex are rotations around three axes and the "tumble of a tennis racket"which was discovered in a bizarre form (Bizarre Behavior) by the Soviet astronaut Dzhanibekov. These tumblings

can arise due to changes in kinetic energy during body rotations, and then again become phenomena that "last to continue to last".

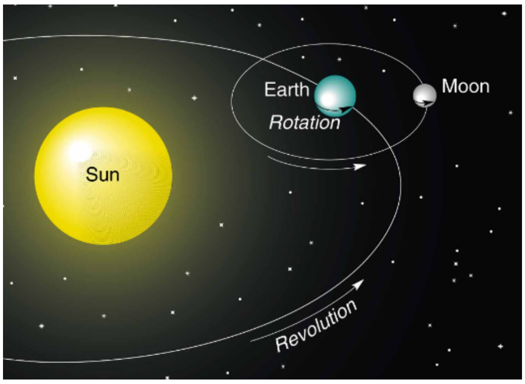
Using Newton's law, we can find the gravitational accelerations $g = GM/r^2$ of the Sun at the level of the Earth's orbit (revolution), the Earth's at the surface of the Earth, or the Earth's at the surface of the Moon, the celestial bodies we see in the following image on the right. Newton's gravitational constant is $G = 6.6743 \times 10^{-11}$ m^3 kg^{-1} s^{-2}, and the masses of the Sun, Earth, and Moon are respectively:

$$M_S = 1.9885 \times 10^{30} \text{ kg}, \quad M_Z = 5.972 \times 10^{24} \text{ kg}, \quad M_M = 7.3477 \times 10^{22} \text{ kg},$$

and the distances of the Earth from the Sun, the radius of the Earth, and the distances of the Moon from the Earth are respectively:

$$r_s = 1.496 \times 10^{11} \text{ m}, \quad r_z = 6.371 \times 10^6 \text{ m}, \quad r_m = 3.844 \times 10^8 \text{ m}.$$

Next, how do we calculate the centrifugal accelerations with which gravity holds these planets in orbits.



The Earth rotates around the Sun with an acceleration of $g_s = 0.059$ m/s^2. The acceleration of the Earth on the ground is $g_0 = 9.81$ m/s^2, and the Moon rotates around the Earth against the Earth's acceleration of $g_m = 0.0027$ m/s^2. Using the formulas (3.18) of the angular momentum $v^2 = ar$, we find the velocities $v = \sqrt{gr}$ of the Earth's motion around the Sun and the Moon's motion around the Earth, respectively:

$$v_z = 29785, \quad v_m = 1018,$$

in meters per second (m/s).

These are some of the average known speeds of the Earth around the Sun (107,000 km/h) and the Moon around the Earth (3,683 km/h) that are more often found using Kepler's third law. However, we started from the expression $v^2 = ar = GM/r$ for the (pseudo) centrifugal force that is in equilibrium with gravity and arrived at the same.

Namely, since $v = 2\pi r/T$ is the speed of orbiting a circle of circumference $2\pi r$ of radius r in time T, it is $T^2 = (4\pi^2/GM)r^3$ and $T^2/r^3 =$ const, i.e. Kepler's third law, with approximately r for the major axes of the ellipses of the orbits of these planets (Earth around the Sun and Moon around the Earth). They are approximately circles. In other words, the ratio of the square of the period of revolution of a planet around the Sun to the cube of the major semi-axis of the orbit is the same for all planets.

As we can see from the attached calculation, and then from the derivation of Kepler's Third Law, this law also applies to the orbit of the Moon around the Earth. However, it also applies to centrifugal motions in general, as we saw in the same article, that is, to angular momentum. That "permanent body connection"that takes away momentum from uniform inertial motion and deflects it into a conic path in the application to the solar system was the gravitational force.

I have often made these various calculations of the same thing in order to verify the accuracy of my calculation, but also to establish the connection of the laws of nature in such a way that each part of the truth can be true both within its own and other correct theories. This is also a kind of (1) observability, in addition to (2) durability and (3) truthfulness.

3.2.3 Time dilation

Interpreting as "real"what can be perceived from what we can perceive, implying for it the kind of consistency (transitivity) that otherwise universal truths have (deductiveness), we here also consider abstract truths to be "real."What is not real is *fictitious* , and both structures of such kinds exist. In this unusual sense, the reality of relativistic time dilation is tested.

First of all, the 4-dim interval of the event (3.15) in the two inertial frames will be equal. From $(ds')^2 = (ds)^2$ it follows:

$$c^2(dt')^2 - (dx')^2 - (dy')^2 - (dz')^2 = c^2(dt)^2 - (dx)^2 - (dy)^2 - (dz)^2,$$

$$c^2(dt')^2 - (d\ell')^2 = c^2(dt)^2 - (d\ell)^2,$$

$$\left(1 - \frac{v'^2}{c^2}\right)(dt')^2 = \left(1 - \frac{v^2}{c^2}\right)(dt)^2.$$

Assuming that the clock is at rest in the first system ($v' = 0$), so it moves with a velocity v relative to the second, it will be:

$$dt = \gamma dt', \quad \gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}. \tag{3.20}$$

This is the time dilation relation (3.10), now derived for differentials. The clock in the moving system lags behind the relative observer by as many times as its durations of seconds are longer. The coefficient of this elongation is Lorentz's γ. In order for the speed of light along that direction of motion to remain the same, the relative observer will see a γ reduction in length units

$$d\ell = d\ell'/\gamma. \tag{3.21}$$

This is formula (3.9) for length contraction, now written for differentials.

Let us apply these infinitesimals to a rotating disk, using formulas (3.18) for the spin. A point rotates in a circle of radius r about its center O with tangential velocity $v = \omega r$, so the circumference of the circle is $2\pi r\sqrt{1 - (\omega r/c)^2}$, decreasing while the diameter remains the same. The circumference divided by the diameter is less than π, and the rotation changes the space into an elliptical, positive *Gaussian curvature*. However, time for distant points flows more slowly, which is why they become attractive. The centrifugal force makes the center of such a space repulsive; it pushes material points away towards the periphery.

A decade or so ago, I considered tensor models of rotating universes (Rotating Universe) including those that are not what I eventually decided on, when they imply a cylindrical rotation and some kind of axis of rotation along the universe. Many such rotations lose their meaning in classical physical 3-dimensional space, but they have their place in the information theory that I was simultaneously and patiently building at the time as a hypothesis that I might eventually be able to challenge. I consider them irrelevant for this presentation.

Partly inverse to the rotations is the gravitational field. For example, in the centrally symmetric gravity of the Schwarzschild metric, time flows more slowly closer to the center, with radial lengths shortening by $\Delta r\sqrt{1-2GM/rc^2}$. A circle circumscribed around the center would have a shortened diameter and unchanged circumference. The quotient of its length and diameter is again greater than π, which redefines 3-dimensional space as a spherical geometry with positive Gaussian curvature. However, the space-time of the event is with an attractive, gravitational force directed towards the center, because now time flows more slowly there.

So, according to the principle of minimalism, systems tend to develop into less informative ones, which means more certain, better ordered, more predictable, therefore leaner free choices and fewer events. Fewer events are the slow flow of time and the attraction that, according to the above, is independent of the contraction of lengths. Contrary to rotating models, for a more promising space-time, the (my) information theory now recommends one that seems to collapse into an increasingly strong gravitational field. If the older lengths are getting larger, here is another[5] confirmation of the ever-faster expansion of the universe.

Example 114. *Let's test the formula $E = mc^2$ on a mass m that is falling freely in a weak (Newtonian) gravitational field, increasing its total energy E by the increase in kinetic energy.*

Solution. We use Newton's law of gravity for a body of mass M and a smaller mass m in free fall straight towards the center of the field[6]:

$$d(mc^2) = -G\frac{Mm}{r^2}dr,$$

$$\frac{dm}{m} = -\frac{G}{c^2}\frac{M}{r^2}dr,$$

$$\ln m = \frac{GM}{c^2r} + \text{const.}$$

$$m = m_0 e^{GM/c^2r},$$

where m_0 would be the mass of the falling body in the absence of gravity. Hence:

$$m = m_0\left(1+\frac{GM}{rc^2}\right) = m_0\sqrt{1+\frac{2GM}{rc^2}},$$

[5]In addition to the more likely transitions of fermions into bosons, more precisely, matter into space, and the unequal aging of vacuum and masses.

[6]From [8], 1.2.8 Vertical fall, now with a different instruction.

by expanding the expression into Maclaurin series and neglecting the higher exponents of small contributions. □

Note that mass is still an undefined property in physics. Physicists disagree on how mass is affected by gravity, and therefore the result of this example and the immediate continuation should be considered speculative, and I present it even though it does not seem to be important for information theory itself. So, in that final expression of the example, we recognize the same square of the velocity $v^2 = GM/r$ used to derive Kepler's third law from the centrifugal force of a rotating disk. Thus:

$$m = m_0\left(1 + \frac{GM}{rc^2}\right) = m_0\left(1 + \frac{v^2}{c^2}\right) = m_0 + \frac{1}{2}\frac{mv^2}{c^2} + \frac{1}{2}\frac{mv^2}{c^2},$$

$$E = E_0 + E_k + \frac{1}{2}m_0v^2,$$

where E_k is the kinetic energy due to the velocity of the falling body, and $\frac{1}{2}m_0v^2 = E_0v^2/2c^2$ is the increase in energy (and therefore mass) of the body at rest due to gravity alone. This would mean that

$$m = \frac{m_0}{\sqrt{1 - \frac{GM}{rc^2}}} \tag{3.22}$$

the relativistic expression for the mass m at rest in a gravitational field at a height r from the center, where m_0 is that mass at rest outside the gravitational field. This also leads to the corresponding expression for the energy

$$E = \frac{E_0}{\sqrt{1 - \frac{GM}{rc^2}}}, \tag{3.23}$$

where E is the energy of the body at rest in the gravitational field, and E_0 is the energy of the same body at rest outside the gravitational field. For bodies in free fall, the fraction under the root should be multiplied by 2 and it becomes a factor with dr in the Schwarzschild metric.

The hypothesis of a larger mass of a body near a gravitational field appeals to me because of the law of large numbers. Statistically, in a larger crowd there is greater certainty, a lower frequency of choices and less uncertainty, and this goes with a slower flow of time. However, this, already because of the mass itself which means inertia and that reduction of options and again greater certainty, also fits with a slower flow of time. Third, the present loses information by building its past and like a truck leaking sand – it becomes thinner, rarefied, and with the above hypothesis we can save some of the missing mass.

Continuing the same hypothesis, we also consider light particles with higher energy near the gravitational field than far outside. But, unlike the simple arrival of light from other sources, this one has to fight its way up, losing some of its energy. We now take this as the objective state of its lower energy outside the field

and, according to the result of the example, light emitted from a distance r from the star will have the frequency:

$$\omega = \omega' e^{-GM/rc^2} \approx \omega' \sqrt{1 - \frac{2GM}{rc^2}},$$

and because $\omega\Delta t = \text{const}$, it will be

$$\Delta t = \frac{\Delta t'}{\sqrt{1 - \frac{2GM}{rc^2}}}, \tag{3.24}$$

where Δt is the period of the observed oscillation from a distance, and $\Delta t'$ is the period of the same oscillation near the star. We recognize this as time dilation (3.20), now by the gravitational field.

The observed light from distant stars climbed out of its gravitational embrace and lost energy, and we continue to take this new state of frequency as objective reality, no longer paying attention to its origin. If time really flows more slowly for an object (3.24), so that its speed of light in the direction of some axis is equal to ours, then its units of length in that direction are that many times shorter. Hence, the unit of length near a star as we see it is

$$\Delta r = \Delta r' \sqrt{1 - \frac{2GM}{rc^2}}, \tag{3.25}$$

where $\Delta r'$ is the intrinsic (local, real) length.

Knowing that *spherical coordinates* are orthogonal and with the source of the central gravitational field at the origin, and assuming that there are no contractions of lengths perpendicular to the radial ones, based on the above we can write an expression for the Schwarzschild event interval

$$ds^2 = \left(1 - \frac{2GM}{rc^2}\right) c^2 dt^2 - \left(1 - \frac{2GM}{rc^2}\right)^{-1} dr^2 - d\Omega^2, \tag{3.26}$$

where $d\Omega^2 = r^2(\sin^2\theta d\varphi^2 + d\theta^2)$ is the element of the sphere perpendicular to the radius. This expression is consistent with (3.15) in the sense that $ds^2 \ge 0$. The speed of light (in a vacuum) is the fastest event of all, so $ds = 0$ only when it represents light.

Example 115. *Time dilation in a gravitational field is*

$$dt = \frac{dt_0}{\sqrt{1 - \frac{2GM}{rc^2}}},$$

where dt_0 is the characteristic time of a point at rest on a sphere $d\Omega = d\Omega'$.

Proof. We observe a point $dr' = 0$ during time dt'. From $ds = ds'$ it follows:

$$c^2 dt^2 - dr^2 = \left(1 - \frac{2GM}{rc^2}\right) c^2 dt'^2 - \left(1 - \frac{2GM}{rc^2}\right)^{-1} dr'^2,$$

$$\left(1 - \frac{dr^2}{c^2 dt^2}\right) c^2 dt^2 = \left(1 - \frac{2GM}{rc^2}\right) c^2 dt'^2,$$

$$\left(1 - \frac{v^2}{c^2}\right) c^2 dt^2 = \left(1 - \frac{2GM}{rc^2}\right) c^2 dt'^2,$$

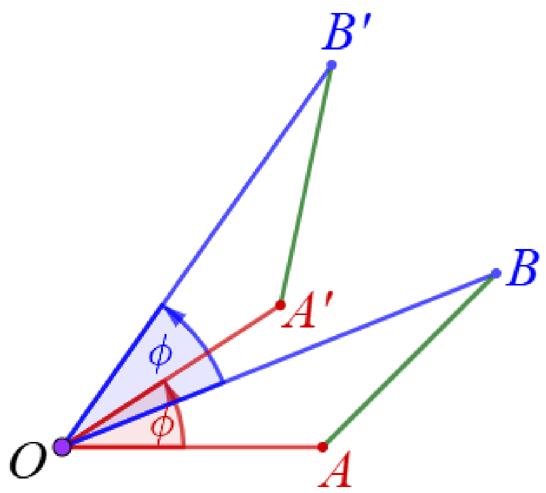
where $dr/dt = v = 0$ is the velocity of the point. Hence

$$dt^2 = \left(1 - \frac{2GM}{rc^2}\right) dt'^2,$$

so, assuming that $dt = dt_0$ is the proper time, and $dt' = dt$ is the observed time, the required equality follows. □

3.2.4 Isometrics

Mappings that do not change the distances of points are called *isometry*. In geometry, these are central, axial, and mirror symmetry, as well as translation and rotation; in algebra, they are unit norm operators; and in (my) information theory, they are "realities." In what follows, I will show that all isometries actually reduce to rotations, and then, in some cases, explain what is meant by "rotation."



First of all, the geometric rotation, $\rho_o(\phi)$ is the mapping given by the point O and the angle ϕ. This is shown in the figure on the left, where $\rho : A \to A'$ and $\rho : B \to B'$, so that:

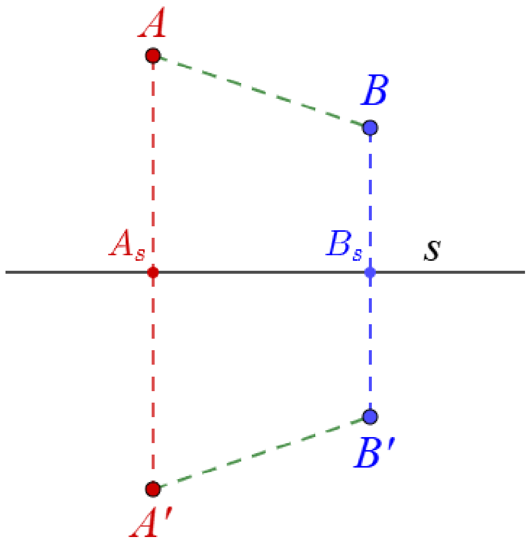
$$\phi = \angle(AOA') = \angle(BOB'),$$

$$OA = OA', \quad OB = OB'.$$

Triangles AOB and $A'OB'$ are congruent, since they have two equal sides and an angle subtended by them, so $AB = A'B'$ and the rotation $\rho_o(\phi)$ is an isometry.

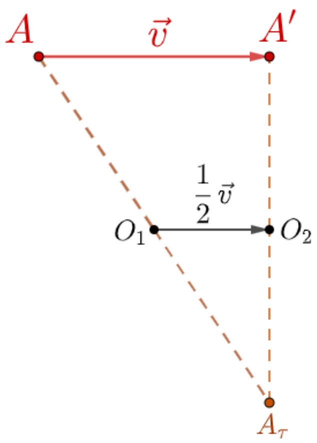
The central symmetry σ_o is given by the point O through which, say, the point A is mapped to A' such that $A - O - A'$, i.e. the three points are collinear (on the same line) with O between A and A', where $AO = OA'$. It is easy to prove that the central symmetry is a rotation by an extended angle, $\sigma_o = \rho_o(180^\circ)$.

In the following figure on the right is the axial symmetry $\sigma_s : AB \to A'B'$ which maps the pair of points A and B to the pair of points A' and B', such that $AA' \perp s$ and $BB' \perp s$, where $AA' \cap s = A_s$ and $BB' \cap s = B_s$, and $AA_s = A_sA'$ and $BB_s = B_sB'$. Hence $A'B' = AB$, so the axial symmetry is an isometry. It maps the triangle into a congruent triangle of the opposite orientation. However, this symmetry is also obtained by rotating about the s-axis in the third dimension by the extended angle.



Mirror symmetry σ_π is analogous to axial symmetry, except that the point A on the plane π is extended to the point A_π in the same direction by the same amount to the copy A'. The result is a mirror image of the original, hence the name of this symmetry. It can also be obtained by rotating in a new dimension about the mirror plane π by an extended angle.

Translation $\tau(\vec{v})$ is a parallel displacement of a point by a given vector $\vec{v}$ when the original is the base and the vertex of the vector is a copy of the given point. In the figure on the left are the central symmetries $\sigma_{o_1} : A \to A_\tau$ and $\sigma_{o_2} : A_\tau \to A'$ which result in a translation $\tau : A \to A'$. Namely, the centroid O_1O_2 of the triangle $AA_\tau A'$ is along (vector $\frac{1}{2}\vec{v}$) the midpoint of two sides, which is parallel to the third and equal to half of it. This is a well-known theorem of geometry that now proves that the vector $\overrightarrow{AA'}$ is equal to the vector $\vec{v}$, i.e. that we have a translation.



As far as elementary geometry is concerned, I hope this is enough to convince us that isometries are represented by rotations. By studying rotations, we study symmetries, translations, and other copying that preserve the distances of points. More broadly, by studying algebraic rotations, we will have the tools to represent the conservation laws of physics, and even more broadly, we will understand reality in terms of information theory.

Algebra simulates these geometric mappings with linear functions. Thus, the central symmetry $\sigma_o(x,y) = (-x,-y)$, the axial symmetry $\sigma_x(x,y) = (x,-y)$, and the plane symmetry $\sigma_\pi(x,y,z) = (x,y,-z)$ are mirrors of the π plane Oxy, i.e. $z = 0$. These are linear operators, and as such they have matrix representations:

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -x \\ -y \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -y \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

These matrices have a determinant of 1, or -1 when changing the orientation of the image. If a matrix does not change the magnitude of a vector, then its determinant is ± 1, but the converse is not true. For example, the matrix:

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x+y \\ x+y \end{pmatrix}$$

has determinant 1 and $x^2 + y^2 \neq (2x + y)^2 + (x + y)^2$, changing the lengths of the copies. To exclude such examples, we use the following definition.

Definition 116. *A bounded linear operator $A \in \mathcal{L}(X)$, an element of the space of linear operators of a vector space X, is an isometry if $\|Av\| = \|v\|$ for all vectors $x \in X$.*

It is easy to show that the rotation matrix is an isometry,

$$\begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x\cos\phi - y\sin\phi \\ x\sin\phi + y\cos\phi \end{pmatrix},$$

because $(x\cos\phi - y\sin\phi)^2 + (x\sin\phi + y\cos\phi)^2 = x^2 + y^2$, which is easy to check, so the determinant of that matrix is one. Let us show that the converse is also true.

Example 117. *In the case of 2-dimensional real vector spaces, scalars of real numbers ($\Phi = \mathbb{R}$), an isometry is a rotation about the origin by some angle ϕ.*

Solution. We see this from the equations $Av = u$ and $\|v\| = \|u\|$:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} v_x \\ v_y \end{pmatrix} = \begin{pmatrix} u_x \\ u_y \end{pmatrix}, \quad v_x^2 + v_y^2 = u_x^2 + u_y^2,$$

$$v_x^2 + v_y^2 = (av_x + bv_y)^2 + (cv_x + dv_y)^2 =$$
$$= (a^2 + c^2)v_x^2 + (ab + cd)v_x v_y + (b^2 + d^2)v_y^2,$$
$$a^2 + c^2 = 1, \quad ab + cd = 0, \quad b^2 + d^2 = 1.$$

We can introduce a change of variables, without loss of generality:

$$a = \cos\alpha, \quad c = \sin\alpha, \quad b = -\sin\beta, \quad d = \cos\beta,$$

when the first and third equality are automatically fulfilled, and the middle one is:

$$\cos\alpha(-\sin\beta) + \sin\alpha\cos\beta = 0,$$

$$\sin(\alpha - \beta) = 0,$$

and hence $\alpha = \beta + k\pi$, $\forall k \in \mathbb{Z}$. The basic solution is $\alpha = \beta = \phi$. □

In the case of Markov chains, for stochastic matrices (whose columns are probability distributions) to be isometries, it is necessary that their determinants are unitary. That is, if the matrix is a permutation of the columns of the unit. Otherwise, if $-1 < \det A < 1$, then $\det A^n = (\det A)^n \to 0$ when $n \to \infty$, so the Markov chain converges to a "black box". Namely, when $\det A = 0$ then the matrix is not invertible.

We often work with complex spaces, the scalar $\Phi = \mathbb{C}$, when the scalar product of the vectors $\langle x, y\rangle = xy^*$. Otherwise, every complex number $x = a + ib$ ($a, b \in \mathbb{R}$) has a conjugate form $x^* = a - ib$, when $x^{**} = x$. The real part of the number x is $\Re(x) = a$, and the imaginary part $\Im(x) = b$. Note that $\Re(x) = \Re(x^*)$, so $\Re(xy^*) = \Re(x^*y)$. Let's apply this to vectors.

Lemma 118. *If $A \in \mathcal{L}(X)$ is an isometry, then $\langle Ax, Ay\rangle = \langle x, y\rangle$ for all vectors $x, y \in X$.*

Proof. Let $x, y \in X$ be given:

$$\|x\|^2 + 2\Re\langle x, y\rangle + \|y\|^2 = \langle x + y, x + y\rangle = \|x + y\|^2 = \|A(x + y)\|^2 =$$

$$= \langle A(x + y), A(x + y)\rangle = \|Ax\|^2 + 2\Re\langle Ax, Ay\rangle + \|Ay\|^2 = \|x\|^2 + 2\Re\langle Ax, Ay\rangle + \|y\|^2.$$

We get $\Re\langle Ax, Ay\rangle = \Re\langle x, y\rangle$. Replacing the vector y with the vector iy in the previous procedure, we also obtain $\Im\langle Ax, Ay\rangle = \Im\langle x, y\rangle$. Therefore, $\langle Ax, Ay\rangle = \langle x, y\rangle$, which means that A is an isometry. □

Unlike physics and other sciences, or engineering, where "we know what is what", in mathematics we easily use the same labels for different concepts, but we have to announce their use. I emphasize this because of this x which is sometimes a vector sometimes a number. There are too many concepts in mathematics for the alphabet; after all, it is thousands of years older than any science.

The message of the previous sections that "what lasts would last", otherwise a paraphrased consequence of the principle of minimalism, is valid through this lemma as well. Not only do isometries preserve information (amounts of freedom) of the subjects themselves, but they also preserve the amounts of their possible communications (connections of information). Simply put, reality cannot be: (1) completely hidden, (2) consumed, nor (3) disputed, i.e. translated into untruth.

Theorem 119. *An operator $A \in \mathcal{L}(X)$ is an isometry if and only if $A^\dagger A = I$, i.e. if its adjoint (conjugate transpose) is equal to its inverse $A^\dagger = A^{-1}$.*

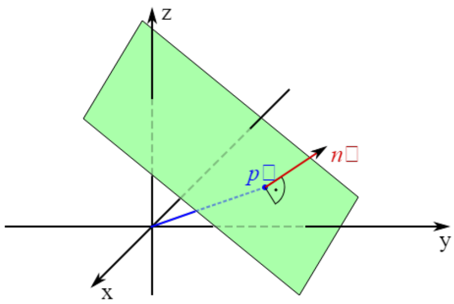
Proof. If $A \in \mathcal{L}(X)$ is an isometry, then $\langle x, y\rangle = \langle Ax, Ay\rangle = \langle A^\dagger Ax, y\rangle$ for all $x, y \in X$. By subtraction, $\langle (I - A^\dagger A)x, y\rangle = 0$ for all $x, y \in X$, so $A^\dagger A = I$.

If $A \in \mathcal{L}(X)$ is such that $A^\dagger A = I$, then for all $x \in X$ we have $\|Ax\|^2 = \langle Ax, Ax\rangle = \langle A^\dagger Ax, x\rangle = \langle x, x\rangle = \|x\|^2$, so A is an isometry. □

The kernel, null space, or null(A) of a linear transformation A is a subset of the domain that transforms into the null vector. Solutions of homogeneous linear systems are an important source of vector null spaces. If A is a matrix m by n, then all vectors $x \in \mathbb{R}$ of the homogeneous system $Ax = 0$ form a single null space.

Another example is in the image on the right. It is a plane in three-dimensional space. It is given by the scalar product of a nonzero vector $\mathbf{n} = (a, b, c)$ and a vector $\mathbf{r}_0$ at a point $P_0(x_0, y_0, z_0)$ with $\mathbf{n} \cdot (\mathbf{r} - \mathbf{r}_0) = 0$, more precisely

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0,$$



with an arbitrary vector $\mathbf{r}$ representing the points of that plane. This is actually the definition of orthogonality, here of the vector $\mathbf{n}$ to the plane passing through the point P_0, that is, to all vectors $\mathbf{r} - \mathbf{r}_0$. This plane is a null-space.

Since no single subject can perceive all objects, because there are always spaces perpendicular to a given vector, then such null spaces are what multiple subjects can observe together. They are observers that have non-zero projections onto such a subspace. For them, partial isometry holds.

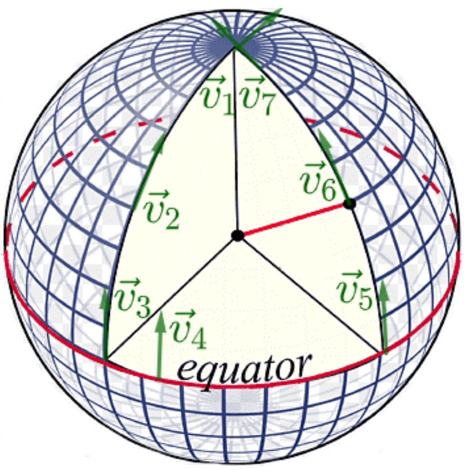
A linear combination of zero-space vectors is a zero-space vector, and subjects of similar perceptions, like individuals of a similar species, can live happily in their communities. However, not perceiving external reality will not negate it, and it may sooner or later return unpleasantly to the colony. Like microbes that we all cannot see without technical aids, say, or radiation, or odorless poisonous gas.

3.2.5 Defect

On the other hand, continuing our consideration of null-space, we know that “subjects of similar perceptions” can in various ways also be producers of phenomena that they themselves cannot perceive. For example, we can feel electricity but not magnetism that arises perpendicular to it. The following diagonal block matrix consists of a matrix A of order m and a matrix B of order $n-m$

$$C = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}.$$

When mapping sequences of length n, the matrix C can be a partial isometry with block A, and something else with block B. The following example is shown in the figure on the left.



The green arrows are the positions of the same vector $\vec{v}_1$, $\vec{v}_2$, ..., $\vec{v}_7$ that is translated from the North Pole along the meridian of the sphere, along the equator, and then back to the North Pole. It is clearly seen that the final vector is not equal to the initial one, $\vec{v}_7 \neq \vec{v}_1$. What if it is a momentum vector or some other physically viable vector quantity?

This phenomenon, the loss of gravitational field energy, was observed shortly after the discovery of the general theory of relativity (1916) and has not yet been explained. I did not consider that this defect should be (only) a leak of energy into the additional dimensions of time, as a colleague may have suggested to me, but rather a leaving of gravitational energy in the current 4-dimensional space-time. According to the theory of relativity, energy is the fourth component of momentum, and according to (my) information theory, the present leaves a trace of the past.

Some of the lost information of the present returns as information of the past, but we suspect that the past also guides future events, denying them some options and, therefore, increasing their certainty. Greater certainty is less uncertainty.

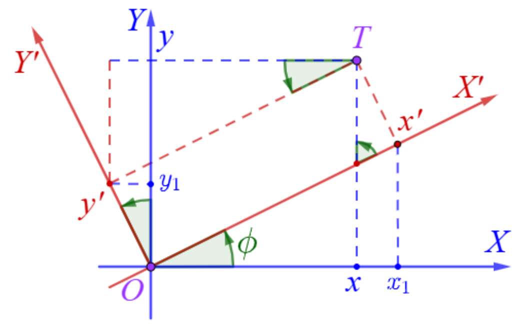
Thinning out information means reducing its saturation. I remind you, an event that we know will happen and it does happen – is not very informative.

If the vector defect after translations along a closed path results in a leak of gravitational energy into the past, and the greater curvature of space would cause this larger anomaly, then more massive bodies would leave a thicker trace. Since *dark matter* is, for now, the only candidate for accepting such a trace, it should be found more in more massive galaxies, in older ones, and on paths that galaxies have crossed more often. I am waiting for astronomy to confirm or refute some of this.

However, by itself bypass can be the cause of "dark matter" and more. Then the metric (3.26) of gravity, or (3.15) of rectilinear uniform motion, is better written

$$ds^2 = dx^2 + dy^2 + dz^2 - c^2 dt^2 \tag{3.27}$$

where again $d\ell^2 = dx^2 + dy^2 + dz^2$ is the square of the distance traveled in time dt. Since the real speed is not greater than the speed of light $v \le c$, or $d\ell/dt \le c$, it is $ds^2 \le 0$, so ds is an imaginary number, except for events on the paths of light when it is zero. Even if we look at our feet, we don't see what is happening now but what was once there, because light takes time to travel the distance from our point of view.



Let us now return to isometries, more precisely to rotations in one plane as in the figure on the right. The rectangular system OXY is rotated in $OX'Y'$ by an angle ϕ. Angles with perpendicular legs are equal (or are supplementary) and are marked as such in green in the figure. The point T has projections onto both systems, respectively (x,y) and (x',y'). In addition, x_1 is the projection of the point x' on the abscissa, and y_1 is the projection of y' on the ordinate.

It is obvious:

$$Ox_1 = Ox + xx_1, \quad Oy_1 = Oy - yy_1,$$

$$x'\cos\phi = x + y'\sin\phi, \quad y'\cos\phi = y - x'\sin\phi.$$

Hence the direct and inverse transformations:

$$\begin{cases} x = x'\cos\phi - y'\sin\phi \\ y = x'\sin\phi + y'\cos\phi \end{cases} \quad \begin{cases} x' = x\cos\phi + y\sin\phi \\ y' = -x\sin\phi + y\cos\phi \end{cases} \tag{3.28}$$

We write the direct ones in matrix form

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix}$$

and these are the ones we found in Example 117, looking for isometries.

On the other hand, continuing (3.27), let us take $r = x$ as the axis of rectilinear inertial motion and $\tau = ict$ the imaginary ($i^2 = -1$) path traveled by light at speed c and simply use (3.28). For direct ones we get:

$$\begin{cases} r = r' \cos\phi - \tau' \sin\phi \\ \tau = r' \sin\phi + \tau' \cos\phi \end{cases} \tag{3.29}$$

so introducing hyperbolic functions[7]:

$$\sinh\phi = \frac{e^{\phi} - e^{\phi}}{2}, \quad \cosh\phi = \frac{e^{\phi} + e^{-\phi}}{2} \tag{3.30}$$

and changing the angle to imaginary $\phi = i\varphi$, i.e. $\varphi = -i\phi$ to have *Euler's formula* $e^{i\varphi} = \cos\varphi + i\sin\varphi$, we find $\sinh\phi = i\sin\phi$ and $\cosh\phi = \cos\phi$, i.e. $\sin\phi = -i\sinh\phi$ and $\cos\phi = \cosh\phi$. Thus (3.29) we write:

$$\begin{cases} r = r' \cosh\phi + ct' \sinh\phi \\ ct = r' \sinh\phi + ct' \cosh\phi \end{cases} \tag{3.31}$$

and these are the Lorentz, or special relativity transformations.

Example 120. *Let us show that (3.31) are the Lorentz transformations*

$$r = \frac{r' + vt'}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad t = \frac{t' + \frac{v}{c^2} r'}{\sqrt{1 - \frac{v^2}{c^2}}},$$

for rectilinear inertial motion with r-axis velocity v.

Proof. Setting the distance traveled $r' = 0$ we find for the velocity of the primed system in the unprimed one:

$$\beta = \frac{v}{c} = \frac{r}{ct} = \frac{\sinh\phi}{\cosh\phi} = \tanh\phi.$$

This is the hyperbolic tangent. However, from (3.30) and this follows:

$$\cosh^2 - \sinh^2 = 1, \quad \tanh^2\phi = \frac{\sinh^2\phi}{\cosh^2\phi},$$

$$(1 - \tanh^2\phi)\sinh^2\phi = \tanh^2\phi, \quad (1 - \tanh^2\phi)\cosh^2\phi = 1,$$

$$\sinh\phi = \frac{\beta}{\sqrt{1 - \beta^2}}, \quad \cosh\phi = \frac{1}{\sqrt{1 - \beta^2}}$$

and by inserting into (3.31) we obtain the Lorentz transformations. □

[7] Wikipedia correct, wrong to correct $\cosh x = \cos x$.

Therefore, Lorentz transformations are one of the isometries, in other words rotations, or expressions of the conservation law, and none of this without walking through the imaginary. It is then supposedly easy to predict that corresponding to these in curved space will also be transformations that are expressions of the conservation law, or rotations, i.e. one of the isometries. Hence it in my book [8] (1.4.2 Rotations).

Given the invariant:

$$ds^2 = (\gamma dr)^2 - (\gamma^{-1} cdt)^2 = (dr')^2 - (cdt')^2, \tag{3.32}$$

which is also valid in a gravitational field when Lorentz's $\gamma = 1/\sqrt{1 - 2GM/rc^2}$ for points at rest in the field, and outside gravity it is $\gamma = 1/\sqrt{1 - v^2/c^2}$, let us introduce an additional coefficient χ as a dimensionless number that determines the initial velocity of a body in free fall in a gravitational field. The number χ defines the height from which a given body started to fall, where $\gamma\chi > 1$.

By switching to infinitesimal lengths dr and dt, where we replace $\cos\phi = \gamma\chi$ and $\sin\phi = \sqrt{1 - \gamma^2\chi^2}$ together with $dr \to \gamma dr$ and $dt \to \gamma^{-1} dt$, we can write the rotation (3.29):

$$\begin{cases} \gamma dr = dr' \cdot \gamma\chi - icdt' \cdot \sqrt{1 - \gamma^2\chi^2} \\ ic\gamma^{-1} dt = dr' \cdot \sqrt{1 - \gamma^2\chi^2} + icdt' \cdot \gamma\chi \end{cases}$$

Dividing the first by γ and the second by γ^{-1}, we get:

$$\begin{cases} dr = \chi\, dr' - \gamma^{-1}\sqrt{1 - \gamma^2\chi^2}\, d\tau' \\ d\tau = \gamma\sqrt{1 - \gamma^2\chi^2}\, dr' + \gamma^2\chi\, d\tau' \end{cases} \tag{3.33}$$

where $\tau = ict$, the imaginary unit $i^2 = -1$, and an arbitrary parameter $\chi \in \mathbb{R}$ is introduced so that the expression under the root is zero or negative. This is so that the "classical coefficient" (generalized Lorentz) remains $\gamma = (1 - 2GM/rc^2)^{-1/2}$, and so that we can also include moving points of the gravitational field.

The formulas can now be checked by calculating:

$$ds^2 = (\gamma dr)^2 + (\gamma^{-1} d\tau)^2 = (dr')^2 + (d\tau')^2, \tag{3.34}$$

which is actually the invariant (3.32) with $d\tau = icdt$ and $d\tau' = icdt'$.

In this way, we reduce the coordinate transformations for the gravitational field to "linear" (on spheres $r =$ const.), and we represent the gravitational field by "rotation", that is, by "isometry". In information theory, by *closed system*, a framework that maintains, say, the totality of energy, we mean the present plus its past. In this sense, (3.33) would be an isometry (without quotes).

The matrix-written $d\mathbf{v} = Ad\mathbf{v}'$ transformations (3.32) are:

$$\begin{pmatrix} dr \\ d\tau \end{pmatrix} = \begin{pmatrix} \chi & -\gamma^{-1}\sqrt{1 - \gamma^2\chi^2} \\ \gamma\sqrt{1 - \gamma^2\chi^2} & \gamma^2\chi \end{pmatrix} \begin{pmatrix} dr' \\ d\tau' \end{pmatrix},$$

with $\det A = 1$, so there is also an inverse $d\mathbf{v}' = A^{-1} d\mathbf{v}$ transformation:

$$\begin{pmatrix} dr' \\ d\tau' \end{pmatrix} = \begin{pmatrix} \gamma^2\chi & \gamma^{-1}\sqrt{1 - \gamma^2\chi^2} \\ -\gamma\sqrt{1 - \gamma^2\chi^2} & \chi \end{pmatrix} \begin{pmatrix} dr \\ d\tau \end{pmatrix}. \tag{3.35}$$

However, by Theorem 119 the operator A is not an isometry, because $A^\dagger \ne A^{-1}$, except when $\gamma = 1$.

Let's do another check of the result, now of the interval (3.34) backwards, using the inverse matrix (3.35):

$$ds^2 = (dr')^2 + (d\tau')^2 = (\gamma^2\chi\, dr + \gamma^{-1}\sqrt{1-\gamma^2\chi^2}\, d\tau)^2 + (-\gamma\sqrt{1-\gamma^2\chi^2}\, dr + \chi\, d\tau)^2 =$$

$$= [\gamma^4 + (-\gamma\sqrt{1-\gamma^2\chi^2})^2]\, dr^2 + [2\gamma^2\gamma^{-1}\sqrt{1-\gamma^2\chi^2} - 2\gamma\chi\sqrt{1-\gamma^2\chi^2}]\, dr d\tau +$$
$$+[\gamma^{-2}(1-\gamma^2\chi^2) + \chi^2]\, d\tau^2 = (\gamma dr)^2 + (\gamma^{-1} d\tau)^2,$$

and this is equality (3.34), that is, (3.32) is exactly where we started.

3.2.6 Harmonizing

Why does the law of conservation apply at all, a colleague once asked me sarcastically, trying to say that everything that "deep science" deals with is "trivial and trifling" (bad, poor) things without an answer to that (according to him, supposedly) essential question. It was in a light discussion about energy. It then occurred to me that "there are no stupid questions, only stupid answers" and that perhaps the idea of information (information is the fabric of space, time, and matter, and uncertainty is its essence) is not wrong, which I then began to develop by looking for its contradiction and, in that case, proving determinism.

The keys to the answer to this "stupid question" lie in the three concepts of "reality," communication, finality, and truthfulness. The first concerns interactions, why there is selectivity in communication, and where the needs and limitations of information transfer come from. The second is a matter of the property of infinity that its real part can be equivalent to the whole, which is not within the reach of perceptions and therefore can count on conservation (say, quantities of energy). The third property of physical reality is that it cannot lie, does not understand lies, and does not look back on them. When something is truly demonstrably false, it does not happen, because what it says is what it thinks and does. To put it bluntly.

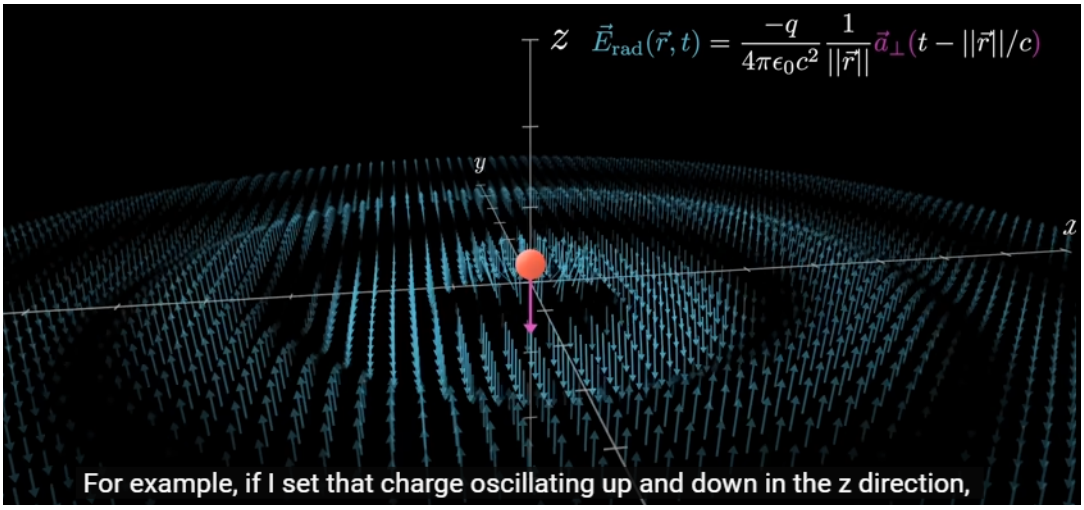
It is a natural phenomenon that more probable outcomes occur more often and less frequent ones are more informative. There is no hunger for information, on the contrary, but there is no extinction (of nature made of information) because there is no infinite divisibility of it. We would renounce all activity, but this death drive is hindered by the finite and therefore indestructible smallest packages of possibilities from which we and everything around us are made. All mathematical statements, steps of proof, legal paragraphs, and all information in general are discrete events; we exchange them in finite portions. Nature can play ping-pong with these quanta of perceptions, in waves or particles, but not drive them into nothingness.

The world we belong to is Berkeleyan in the sense of his slogan that "to be is to be perceived," not because it is the only possible one but because I am not in the other, or we are not possible there. On the other hand, there is the "force of attraction" that I can paraphrase with the dilemma of Euthyphro from Plato's dialogue

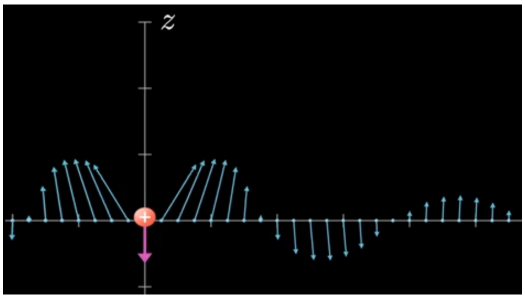
in which Socrates asks him, “Is the pious (ὅσιος– worthy of respect) loved by the gods because he is pious, or is he pious because the gods love him?” The person in question does not want to stop being pious, and the gods will continue to love him. That is the thing with this environment that wants to get rid of information, and without information it cannot. I am not saying that it is easy to understand.

A similar attraction occurs every time with the principle of least action in general or with light covering distances by the shortest paths or in the shortest times, according to its limitations. If these speeds were infinite, little would remain of the law of conservation. If their realizations (truths) were fleeting, little would remain of all nature; the same is true if they did not last. It is irrelevant for the mathematical exposition whether we see the connection of these three concepts (perceptibility, sustainability, and truth), but it is useful; it is inspiring in the search for broader new answers.

On the other hand, similar repulsions occur with the action and reaction of forces in ordinary collisions and even the induction of a magnetic field by an electric one and vice versa. The transfer of the action by the external product of the vectors $\vec{a}$ and $\vec{b}$ to the vector $\vec{c} = \vec{a} \times \vec{b}$ is proportional to the intensity of both factors ($|\vec{a}|$ and $|\vec{b}|$), but as if that were not enough for it, it is also directed perpendicular to both vectors ($\vec{c} \perp \vec{a}$ and $\vec{c} \perp \vec{b}$) as if there were no more information of perception with them ($\vec{c} \cdot \vec{a} = 0$ and $\vec{c} \cdot \vec{b} = 0$). However, the overall energy of that and the other part is constant, and this consistency is the mainstay of logic in the investigation of reality. Through the paths of accurate deductions, truth –> truth, reality is traced, not to say captured by them.



The image shows the oscillation of an electric charge along the z-axis (Optics puzzles 2), which creates electromagnetic waves, which do nothing by themselves until an opportunity arises to transmit information from the host (impulse, spin). However, the latter also works on the same fishing for victims, so their effect is illusory.



In the image on the left, we see the same light spreading along one axis to make it easier to see the perpendicularity of the field vector to the direction from the charge point (in the Oxy plane), which moves up and down (along the z-axis), and these oblique arrows to the direction of motion of the (still virtual) photons. This electric field is not perpendicular to the momentum but is tilted slightly in the direction of the momentum. This is a novelty[8], an important note for a consistent interpretation of the motion of light waves using Maxwell's equations.

While this slope exists, there is a "commutator surface," the outer product of the vectors $\nabla \cdot \mathbf{E} = \rho/\epsilon_0$, the first factor of which (nabla) represents the momentum. The further away from the charge, the field slope decreases (the arrow in the figure straightens) and the field weakens. Its amplitude and density ρ decrease, but not the wavelength λ or the frequency $f = c/\lambda$ of light. The chances of an interaction by which the source of the wave, the information, could carry its load to the eventual reader also decrease. The squares of the amplitude modulus are the interaction probabilities.

Even weaker amplitudes can participate in some action and then always transmit the same impulse (the same λ). For example, the impulse of red light with a wavelength of about 700 nm is:

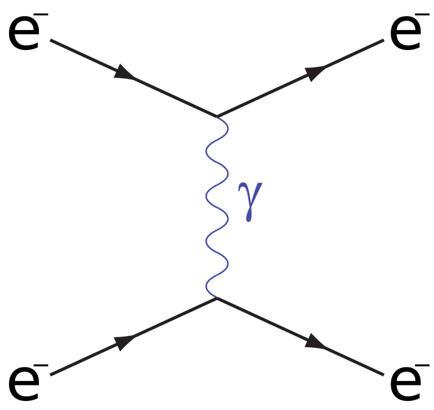
$$p = \frac{h}{\lambda} = \frac{6.63 \times 10^{-34}\ \mathrm{J \cdot s}}{7.00 \times 10^{-7}\ \mathrm{m}} = 9.47 \times 10^{-28}\ \mathrm{J \cdot s/m}.$$

The units of joule-seconds per meter (J·s/m) are equivalent to the units of kilogram-meters per second (kg · m/s), so the momentum of a photon is 9.47×10^{-28} kg · m/s and that is a very small value, but to him, even a little relief would mean a lot.

But alas, when an interaction occurs and the photon is no longer virtual but real, the first one receives as much momentum decrease as the second one increases, in the same direction and opposite directions, so that the total amount is unchanged. The electrons then move away. The spin of light, for example +1, can be subtracted from the sender electron of spin +1/2 and added to the receiver of spin -1/2. After the interaction, the first one has spin -1/2 and the second one +1/2 so that the sum remains unchanged. Thus, the actions are carried through space and time with *alignment* thanks to various chances, principled minimalism, and the law of conservation.

This is nicely represented by the Feynman's diagram in the image on the right. However, how to understand the simultaneity of the transfer of all information of a virtual photon in this transfer, or how they do not carry parts of the electron, thereby consuming it, and similar questions are not elements of the story of Feynman diagrams, nor does physics today have good answers to them.

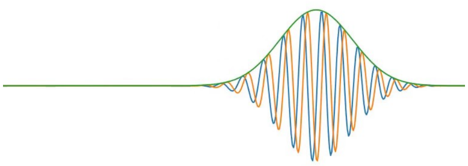
[8] I wrote in section "3.2.1 Light" after formula (3.17).



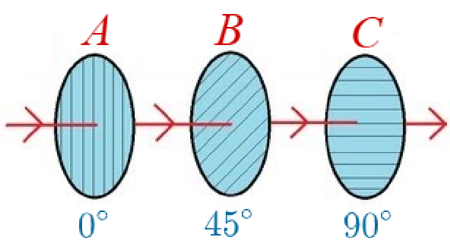
However, with its understanding of the present and its connections with the past, information theory can add something to this. First of all, the fact that a "virtual photon" is not a material point but a "virtual sphere" whose unit area loses density with the square of the distance from the center, and all places are "simultaneous" to it.

By losing density, the virtual sphere loses amplitude, not wavelength, and thus the chance for interaction during which the delivery would always be of the same magnitude. The simultaneity of parts of the sphere's surface means the absence of mutual information transfers of its parts by eventual delivery to another particle, I believe[9], known as *quantum entanglement*.

A charge with its field forms a single whole. For example, it is not possible for two simultaneous interactions to occur so that an electron gives up its spin twice by +1. They can only exchange their energy one at a time, no matter how spatially distant the possible events are at a given moment. On the other hand, these interactions are also quantized, so one must be careful with the concept of a "single event."



Wave phenomena seem to lead a double life, one inside and one outside, which makes the micro-world of physics particularly remote from our everyday life. Take, for example, the wave packet, pictured on the left, which can simulate a series of people rhythmically raising and lowering their arms. When such a group of people moves, the wave packet moves faster than them, but the delivery they would bring arrives only when the entire group, the first person in the group, arrives.



The same is true of a quantum particle-wave. It arrives only when the particle arrives, although the wave of the particle itself may occur faster (in the direction of the packet) or slower (against its movement). There is no delivery of pieces of information until the entire package arrives.

In the picture on the right is the paradox three polarizers of light. Some crystals have molecules arranged in such a way that they only transmit light of vertical waves; in the picture on the right is filter A. When such a filter is rotated by $45°$, filter B

[9] My personal opinion, and I am not aware of it existing anywhere else.

is obtained, and if it is rotated by another $45°$, filter C is obtained. Ordinary light is polarized in all directions, so passing it through filter A, we get only 50 percent of it. Passing that light through A and then through C will not get anything out of it. However, if we put filter B in between, about 12.5 percent of the initial light appears at the output. Paradoxically, two filters (AC) stop more light than three filters (ABC). Such experiments reveal to us the statistical nature of waves, that a wave packet is actually a packet of beliefs.

This "double life" of waves and their ability to simulate "particles" manages to completely fool us. First of all, so much so that it is so difficult for us to distinguish information from energy or action (the product of the change in energy for a given time) and to notice "only" the equivalence between them. We see the latter from the property of belonging of sets, $A \subseteq B \wedge B \subseteq A \iff A = B$. Here it means only the equivalence of sets A and B, the possibility of bijection (mutually unambiguous mapping), i.e. the equal number of their elements, because we say that there is no action without the transfer of some physical information, nor is there a transfer of (physical) information without an action.

We cannot further unpack the smallest waves into their elements, mere information, separate them, and measure them separately using classical methods of physics, but we can still indirectly dissect their components like Archimedes when he found the composition of the crown of King Hiero II. By helping the abstract with the concrete.

3.3 Experiments

In quantum mechanics, an *observable* is any property that can be observed, something that can be measured in a laboratory and for which we have a corresponding operator. For example, the Hamiltonian, $\hat{H} = \hat{T} + \hat{V}$, which measures the kinetic plus potential energy of a particle, also observables $\hat{T}$ and $\hat{V}$. The kinetic energy is the operator

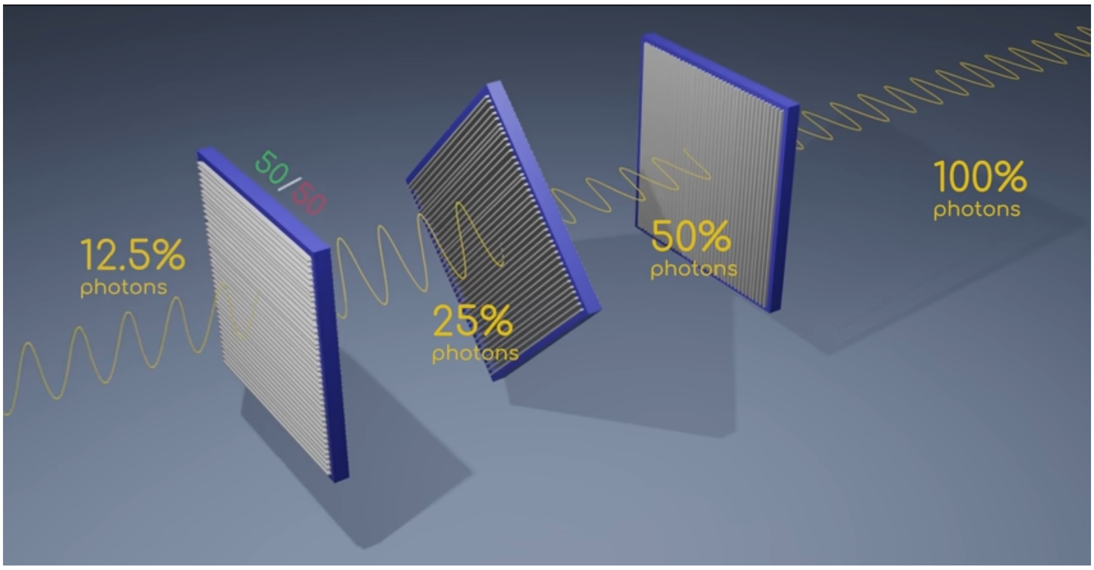
$$\hat{T} = -\frac{\hbar^2}{2m} \cdot \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right),$$

in the Cartesian rectangular system $Oxyz$, and the potential $\hat{V}$ depends on the conditions. Every quantum system has a Schrödinger equation where we apply the Hamiltonian operator $\hat{H}\psi = E\psi$ to the wave function ψ and as a result we get the energy E of the system. In particular, an observable is information, or in classical physics, an observable is an observable physical property or physical quantity that can be measured.

3.3.1 Polarization

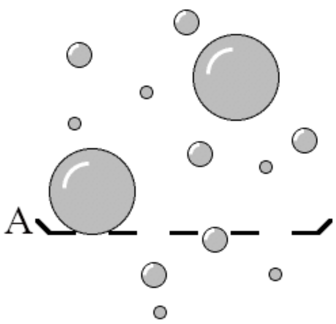
The three-polarizer paradox is a popular light experiment originally described by theoretical physicist Paul Dirac in his 1930 book Principles of Quantum Mechanics. Half (50%) of the light passes through the first filter (View of Light), half

(25%) through the second, and half (12.5%) through the third. It is a probability calculation.

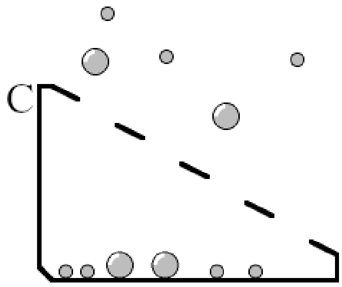


The paradox arises from the counterintuitive result that light can still pass through the third polarizer even though it is perpendicular to the first polarizer, after passing through the middle nugget at 45° on both. Light would not pass through the first and third polarizers, which are oriented at 0 and 90 degrees, respectively. However, the presence of the third polarizer completely changes the situation, as it partially affects the orientation of the light (Material Explanation).

One reason for the eerie nature of these results is the use of the word "filter." A filter is usually understood to mean a device that removes some items from the flow while leaving others essentially intact. An example of a filter is a sieve that blocks objects of a certain size while allowing objects of other sizes to pass through. The following two images, on the left and right, show sieve A, which only lets small beads through, and sieve C, which only lets large ones through. Let's use them to try to understand the meaning of the "three polarizers" paradox and its resolution.

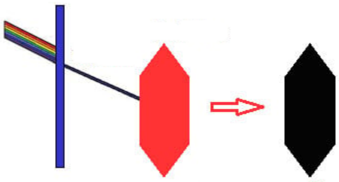


In the picture on the left is a sieve A that passes small and medium-sized balls, but stops large ones. By choosing the size of the lower openings according to the size of the balls, when there are many different sizes, we can regulate the percentage of those that pass and, for example, achieve that exactly 50 percent of them pass. When the largest size of the passed ball is known, in the case of precisely measured, equal and rigid sieve openings, there is no coincidence. There will not be a single ball in the passes larger than the permitted size, and if after A the passed balls are directed to the next filter C, with the same openings as A, not a single ball will pass through it.



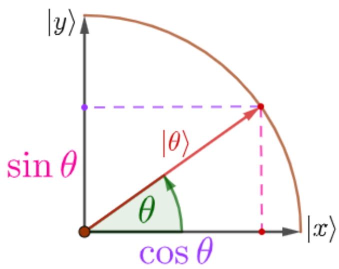
In the picture on the right is a sieve that stops small and medium-sized balls in the lower container, and lets large ones go down the slope. If the openings on the slope of this filter, C, are precisely measured, of equal and rigid openings, adjusted so that only those balls that could not pass through the previous filter A can pass through, then after A and then C, none of the balls will pass through.

Now let us imagine that between these filters we put a third filter B which could moisten the balls with something so that they swell and increase in size. Then some of those behind A, after being moistened through B, could still survive C. A similar thing would happen without filter B, if the openings of the sieve were not quite precisely measured, equal and rigid openings, that is, if there were a little bit of randomness. However, in the micro world there is enough of this necessary randomness.



With light filters, this is not always the case (imprecise, uneven, and loose). We know this from the absorption of colored light (Absorption). The image on the left shows (white) light that is blocked by a blue filter that blocks all colors except blue. That's why that filter looks blue. Then, that blue light encounters a red filter that blocks everything except red. There is no light at the output (black). So, two filters can block all light, and with the third in between, light appears at the output. Let's see what quantum mechanics has to say about this.

A photon polarized at an angle θ to the vertical can be written as a linear combination (superposition, let's say distribution) of a horizontally polarized photon $|x\rangle$ and a vertically polarized photon $|y\rangle$. When $|x\rangle$ and $|y\rangle$ are the ground states of polarization, this means $\langle x|x\rangle = \langle y|y\rangle = 1$ and $\langle x|y\rangle = \langle y|x\rangle = 0$.



In the figure on the right we see the first quadrant of the unit (trigonometric) circle with the mutually perpendicular unit ground state vectors, $|x\rangle$ and $|y\rangle$, which are here on the coordinate axes. In between is the vector $|\theta\rangle$ with the projections $\langle x|\theta\rangle = \cos\theta$ and $\langle y|\theta\rangle = \sin\theta$ onto them. This gives the vector equality $|\theta\rangle = |x\rangle\langle x|\theta\rangle + |y\rangle\langle y|\theta\rangle$.

The probability that a photon polarized at an angle θ will pass through a horizontal polarizer is $|\langle x|\theta\rangle|^2 = \cos^2\theta$, and that it will pass through a vertical one is $|\langle y|\theta\rangle|^2 = \cos^2\theta$. That this is indeed a distribution (superposition) follows from the non-negativity of these probabilities and their unity sum.

The light that falls on the first polarizer $|x\rangle$ is unpolarized, but the photons that pass through the horizontal polarizer are horizontally polarized, so their output is $\theta = 0°$. If they encounter the same horizontal polarizer again, the probability of passing through is $|\langle x|x\rangle| = \cos^2 0° = 1$ and all the arrivals pass through. However, if

they encounter a vertical polarizer, the probability of passing through is $|\langle y|x\rangle|^2 = \cos^2 90^\circ = 0$ and none of the arrivals go further.

Let us now write a horizontally polarized photon as a linear superposition of some other ground states, for example $|+45^\circ\rangle$ and $|-45^\circ\rangle$. Then:

$$|x\rangle = |+45^\circ\rangle\langle +45^\circ|x\rangle + |-45^\circ\rangle\langle -45^\circ|x\rangle =$$

$$= |+45^\circ\rangle\cos(+45^\circ) + |-45^\circ\rangle\cos(-45^\circ) = \frac{1}{\sqrt{2}}|+45^\circ\rangle + \frac{1}{\sqrt{2}}|-45^\circ\rangle.$$

When we insert one of the $\pm 45^\circ$ polarizers between the previous horizontal and vertical ones, the photons will pass through the vertical polarizer, here's how.

With probability $|\langle 45^\circ|x\rangle|^2 = \cos^2(45^\circ) = 0.5$ a horizontally polarized photon passes through a polarizer oriented 45°. Then the photon is in the state $|45^\circ\rangle$, and due to superposition the photon in this state is a linear combination of $|x\rangle$ and $|y\rangle$, i.e.:

$$|45^\circ\rangle = |x\rangle\langle x|45^\circ\rangle + |y\rangle\langle y|45^\circ\rangle = |x\rangle\cos(45^\circ) + |y\rangle\sin(45^\circ).$$

Therefore, $|\langle y|45^\circ\rangle|^2 = \sin^2(45^\circ) = 0.5$ is the probability of a photon passing through a vertical polarizer. Additionally, the probability that a photon exiting a horizontal polarizer will pass through the third vertical polarizer after the middle 45° polarizer is calculated as follows:

$$|\langle y|45^\circ\rangle\langle 45^\circ|x\rangle|^2 = |\sin(45^\circ)\cos(45^\circ)|^2 = \frac{1}{4},$$

and with the 50 percent passed through the first filter, we get the final 12.5 percent (one eighth of all), as the experiment shows.

Something like "looking around the corner"is happening here, or bypassing. The states $|x\rangle$ and $|y\rangle$ do not communicate with each other, because they do not have a scalar product, $\langle x|y\rangle = 0$, but each of them exchanges messages with $|45^\circ\rangle$, because $\langle x|45^\circ\rangle = \langle 45^\circ|y\rangle = 1/\sqrt{2}$, so what passes through $|x\rangle$ can interact with $|45^\circ\rangle$, and some of what is filtered out can also pass through the filter $|y\rangle$.

The unpolarized[10] light illuminating the first polarizer is a 50-50 mixture of any orthogonal set of polarizer angles from -0° to 90°, or from -45° to 45°, or in general from θ to $\theta + 90^\circ$. Otherwise, we use radians for angles ($360^\circ = 2\pi$ radians). The mixture cannot be represented by a single wave function, so for a mixture of unpolarized photons the probability of passing through a vertical polarizer is calculated as follows:

$$\langle\hat{A}\rangle = \sum_k p_k\langle\psi_k|\hat{A}|\psi_k\rangle \tag{3.36}$$

where $\hat{A}$ represents the operator associated with the vertical polarizer $|v\rangle$, and p_k is the coefficient of the photon's parts (probabilities) in the state $|\psi_k\rangle$. In general, for a beam of unpolarized light incident on a vertical polarizer, the expected value for the transmission is:

$$\langle\hat{A}\rangle = \frac{1}{2}\langle\theta|\hat{A}|\theta\rangle + \frac{1}{2}\langle\theta + \frac{\pi}{2}|\hat{A}|\theta + \frac{\pi}{2}\rangle = \frac{1}{2}(\cos^2\theta + \sin^2\theta) = \frac{1}{2}.$$

[10] Feynman, R. P.; Leighton, R. B.; Sands, M. The Feynman Lectures on Physics, Vol. 3; Addison-Wesley: Reading, 1965

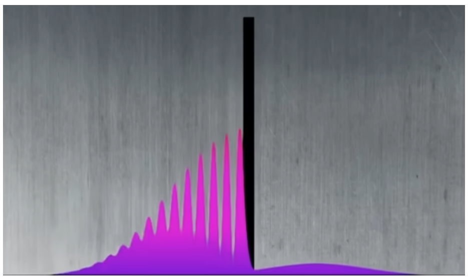
The above strange "bypassing"phenomena observed with polarization of light also occur in the "tunneling effect"of quantum mechanics.

3.3.2 Tunneling

The tunneling effect or tunnelling is a phenomenon in which an atomic particle can overcome a finite potential barrier even when its energy is lower than the height (energy) of the barrier. According to classical physics, this would be impossible; however, according to the laws of quantum mechanics, it is possible.

The author of the video in the image on the left (Tunnel In Real) has an interesting way to explain the “tunnel effect” using a glass of water. Looking at the glass from above, part of it is transparent up to the surface of the water but not below. The water refracts the light in such a way that the walls of the glass become opaque. However, if we press the glass, we can see fingerprints from above where the glass surface and fingers touch. Light penetrates, similar to electrons, through a voltage that they “should not pass through,” but they do pass through in the case of quantum tunneling.

The tunnel effect was first observed experimentally by Robert Williams Wood in 1897, observing the motion of electrons in an emission field, but he was unable to explain it. As early as 1899, researchers in the field of radioactive decay had expressed vague doubts about the possibility that decay occurred due to the tunnel effect, which was not described theoretically until George Gamow in 1929, after the earlier discovery by Rutherford and his colleagues that the alpha particle was actually a helium nucleus. Although the discovery of the tunnel effect is attributed to Gamow (who named it that way), the first theoretical description was given in 1926/27 by Friedrich Hund to describe isomerism in molecules.



As shown in the image on the right, a quantum particle is a probability wave that has some chance of being found outside the barrier, Sabine Hossenfelder explains in her video (About Tunneling). The probability of transmitting a wave packet through a barrier decreases exponentially with the height of the barrier, the width of the barrier, and the mass of the tunneling particle, so tunneling is most pronounced in low-mass particles such as electrons or protons that tunnel through microscopically narrow barriers.

Electron tunneling is easily observed with barriers up to about 1–3 nm thick and up to 0.1 nm for heavier particles such as protons or hydrogen atoms. These experiments are predicted with extraordinary accuracy by the Schrödinger wave equation of quantum mechanics, but although its verification is not an easy task,

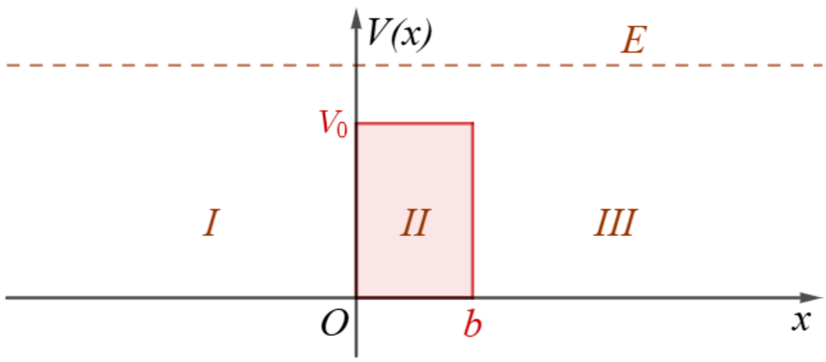
they do happen and accumulate. Over time, very low-probability tunneling experiments with much larger objects have become available, which also confirm the expectations of Schrödinger's probabilities.

One such recent project (Very Slow Reaction, 2023) on the slowest reaction of charged particles ever observed lasted 15 years. These chemical reactions were only theoretical, so when asked whether they could be achieved in real-world experiments, Dr. Roland Wester and the study's author answered positively: "This requires an experiment that allows very precise measurements and that can still be described quantum mechanically."Already in 2018, it was calculated that its quantum tunneling occurs only once in every 100 billion collisions, and the results of this latest study not only confirm those calculations but also promise the development of a precise theoretical model of the tunneling effect.

"Quantum mechanics allows particles to break through an energy barrier due to their quantum mechanical wave properties, and a reaction occurs,"explains Dr. Robert Wild, who is a postdoctoral researcher at the University of Innsbruck and also an author of the study. "In our experiment, we trap possible reactions for about 15 minutes and then determine the amount of hydrogen ions formed. By counting, we find out how often the reaction occurred."

In the book [10] (1.4.6 Potential obstacles), you will find more detailed solutions to problems involving the passage of particles through barriers, and here we will look at just two typical situations. The following figure shows a diagram with the abscissa x along which a particle-wave of energy E moves. On its path, in the interval from 0 to b, there is a potential $V_0 < E$, which is a surmountable obstacle for it given its higher energy. The ordinate is the potential $V(x)$.

Classically, the particle crosses the barrier each time, but its velocity decreases when passing through the potential V_0, because the barrier reduces its kinetic energy, $E_k = \frac{1}{2}mv^2 = E - V_0$, as the particle penetrates through that region II. The momentum of the particle is $p = mv$, so $p^2 = m^2v^2 = 2mE_k$, or $p = \sqrt{2mE_k}$.



Example 121. *Let's solve the Schrödinger equation for the case of the potential barrier from the image.*

Solution. We consider three regions of the potential state for a particle-wave:

$$V(x) = \begin{cases} 0, & x \le 0, \\ V_0, & x \in (0, b), \\ 0, & x \ge b. \end{cases}$$

With the same separation we have the momenta:

$$p(x) = \begin{cases} p_1 = \sqrt{2mE}, & x \le 0, \\ p_2 = \sqrt{2m(E - V_0)}, & x \in (0, b), \\ p_3 = \sqrt{2mE}, & x \ge b, \end{cases}$$

and the wavenumbers by zones $k_1 = p_1/\hbar$, $k_2 = p_2/\hbar$ and $k_3 = p_3/\hbar$. Then we write the time-independent Schrödinger equations separately for each region:

$$\frac{d^2\psi_1}{dx^2} + \frac{2m}{\hbar^2} E\psi_1 = 0 \quad \Rightarrow \quad \psi_1'' + k_1^2\psi_1 = 0,$$

$$\frac{d^2\psi_2}{dx^2} + \frac{2m}{\hbar^2}(E - V_0)\psi_2 = 0 \quad \Rightarrow \quad \psi_2'' + k_2^2\psi_2 = 0,$$

$$\frac{d^2\psi_3}{dx^2} + \frac{2m}{\hbar^2} E\psi_3 = 0 \quad \Rightarrow \quad \psi_1'' + k_3^2\psi_3 = 0.$$

The wave functions are solutions of the form:

$$\psi_1 = A_1 e^{ik_1x} + B_1 e^{-ik_1x},$$

$$\psi_2 = A_2 e^{ik_2x} + B_2 e^{-ik_2x},$$

$$\psi_3 = A_3 e^{ik_1x} + B_3 e^{-ik_1x},$$

where we have taken into account $k_1 = k_3 = p_1/\hbar$, and for the imaginary unit it is $i^2 = -1$. Each solution represents a time-independent component of the traveling wave, but if we consider the I-zone wave function and include the time dependence via $\Psi_1 = \psi_1 e^{-i\omega t}$, we get

$$\Psi_1 = A_1 e^{i(k_1x - \omega t)} + B_1 e^{i(k_1x + \omega t)}.$$

Each of the six coefficients (A_j and B_j, $j = 1, 2, 3$) refers to the amplitudes. The first sum in all three functions represents the wave's movement to the right, and the second, with a minus in the exponent, represents the movement to the left. In this case, the third function has $B_3 e^{-ik_1x} = 0$, because the wave just leaves without returning. □

By setting the initial conditions to the above solutions, for example:

$$\begin{array}{ll} \psi_1(x = 0) = \psi_2(x = 0) & \psi_1'(x = 0) = \psi_2'(x = 0) \\ \psi_2(x = b) = \psi_3(x = b) & \psi_2'(x = b) = \psi_3'(x = b) \end{array}$$

we set a system of equations for a closer determination of the coefficients:

$$A_1 + B_1 = A_2 + B_2, \quad ik_1A_1 - ik_1B_1 = ik_2A_2 - ik_2B_2,$$

$$A_2e^{ik_2b} + B_2e^{-ik_2b} = A_3e^{ik_1b}, \quad ik_2A_2e^{ik_2b} - ik_2B_2e^{-ik_2b} = ik_1A_3e^{ik_1b}.$$

With them we define the probability of transmission and reflection, respectively:

$$P_T = \frac{|A_3|^2}{|A_1|^2}, \quad P_R = \frac{|B_1|^2}{|A_1|^2}. \tag{3.37}$$

The first is the probability that the particle-wave will pass the potential barrier, the second is the probability that it will be reflected from it, that it will not pass, such that $P_T + P_R = 1$.

By eliminating B_1 from the first two equations and eliminating B_2 by adding the third and fourth, and then adding them to eliminate A_2, we obtain equations from which we can obtain

$$\frac{A_1}{A_3} = \frac{e^{ik_1b}}{4k_1k_2}[(k_1+k_2)^2e^{-ik_2b} - (k_1-k_2)^2e^{ik_2b}].$$

Using Euler's formulas for complex numbers:

$$\cos\phi = \frac{e^{i\phi} + e^{-i\phi}}{2}, \quad \sin\phi = \frac{e^{i\phi} - e^{-i\phi}}{2},$$

after calculating and arranging, we find the penetration coefficient

$$P_T = \frac{|A_3|^2}{|A_1|^2} = \left(1 + \frac{V_0^2\sin^2(k_2b)}{4E(E-V_0)}\right)^{-1}, \quad E > V_0, \tag{3.38}$$

for a particle-wave energy greater than the potential barrier.

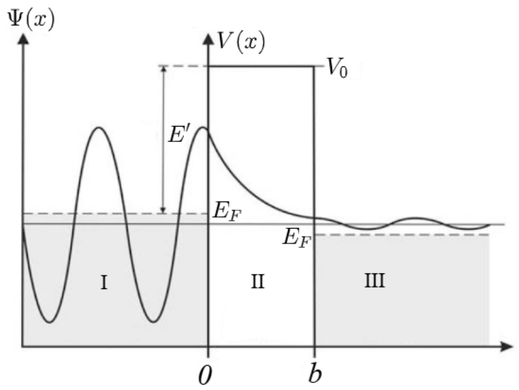
This formula, for transmission in the case of $E > V_0$, predicts two different outcomes. First, when the sine is zero ($k_2b = n\pi$, for $n = 1, 2, \dots$) and this coefficient is 1, which means certainty of passage, the result looks like the classical outcome where the exact number of wavelengths fits into the width of the potential barrier. Second, in all other cases, when this coefficient is less than 1, which means the chances of not passing the barrier, then the reflection coefficient $|B_1|^2/|A_1|^2 > 1$. This means that there is a positive probability that the particle will be reflected even when it has enough energy to overcome the barrier.

Before we continue with particle energies smaller than the potential barrier, let us recall Gödel's theorem on incompleteness, that a consistent theory (of truths and only truths) is not complete (it does not have all the truths). Mathematics also needs untruths to arrive at truths. This is an advantage over consistent theories, but it distances us from physical reality, which does not know how to lie, does not perceive lies, and does not react to them. We see something similar here with "bypassing,"or, to be more precise – with complex numbers.

Without the use of complex numbers, we sometimes do not have access to the reals we need to describe physical reality. Just as deductions that only follow truths

along a consistent network of implications, truth –> truth –> ..., so physical reality remains incomplete without the occasional jump using complex numbers to some other network of implications. Objective chance gives it the possibility for these črazy"jumps into the world of living beings (larger choices) using untruths, or into further truths from networks of one deduction to another, or from one physical reality to another using imaginary numbers.

Tunneling proves to us that such jumps are not breaks in reality. Then there is communication with the "other shore." The law of conservation applies to both sides, and we establish that both are ultimately true. Unlike observable, permanent, and true shores, the very "rivers" that separate them can also be fictions, not consistently observable, and moreover, unstable and untrue.



Classically, a particle of energy $E < V_0$ is always reflected ($P_R = 1$) from a potential barrier. However, particles described by the quantum wave function Ψ, when they lack some of the E' energy to the level of the potential barrier II in the figure on the left, have a small but finite possibility to "tunnel"through the barrier, to pass from region I and continue on the other side to region III. Here[11] is E_F the Fermi level of a metal electrode, otherwise a solid body, which is the thermodynamic work of adding one electron to the body.

For this situation ($E < V_0$) we can use the previous results ($E > V_0$) with the following:

$$k_2 = \frac{\sqrt{2m(E - V_0)}}{\hbar} = \frac{\sqrt{-2m(V_0 - E)}}{\hbar}$$

is an imaginary number. Therefore, we define:

$$\kappa = ik_2 = \frac{\sqrt{2m(V_0 - E)}}{\hbar} \in \mathbb{R},$$

so that the wave function inside region II can be written:

$$\psi_2 = A_2 e^{ik_2 x} + B_2 e^{-ik_2 x} = A_2 e^{\kappa x} + B_2 e^{-\kappa x}.$$

The wave function inside the barrier now becomes exponential, instead of oscillatory. When calculating the probability of transmission and reflection for $E < V_0$, this shift now leads to:

$$\frac{A_1}{A_3} = \frac{e^{ik_1 b}}{4k_1\kappa}[4k_1\kappa\cosh(\kappa b) + 2(k_1^2 - \kappa^2)\sinh(\kappa b)],$$

[11] Frank Trixler: Quantum Tunnelling to the Origin and Evolution of Life; Current Organic Chemistry, 2013

where hyperbolic cosine and sine are used:

$$\cosh x = \frac{e^x + e^{-x}}{2}, \quad \sinh x = \frac{e^x - e^{-x}}{2}.$$

By combining and arranging the expressions, we arrive at the transmission probability:

$$P_T = \left|\frac{A_3}{A_1}\right|^2 = \left(1 + \frac{V_0^2 \sinh^2(\kappa b)}{4E(V_0 - E)}\right)^{-1}, \quad E < V_0. \tag{3.39}$$

Since $\kappa b > 0$, $P_T > 0$ and there is some chance that the particle will pass through the barrier and continue on the other side.

The microworld is dominated by coincidences that, with increasing complexity and in accordance with the law of large numbers of probability theory, slowly transition into the certainties of the macroworld. Then the transmission coefficient $P_T \to 0$ and the reflection coefficient $P_R \to 1$. However, we cannot say that, simply because they have more choices, elementary particles have more vitality.

Also, according to the information of perception, the book [7], the ability to use opportunities (a) is directly proportional to the freedom of action (s) and inversely proportional to the amount of constraints (b). Therefore, $a = s/b$, and hence the freedom of a (living or non-living) system is $s = ab$. If we consider this as a single sum in a series of products of $n = 1, 2, ...$ independent observations, we have a total of $S = s_1 + s_2 + ... + s_n$ freedom of action, or

$$S = a_1 b_1 + a_2 b_2 + ... + a_n b_n.$$

These sums tell us about the possibilities of individual observables, and as we see, they grow not only with possibilities (a_k) but also with impossibilities (b_k, where $k = 1, 2, ..., n$). Order is as necessary to action as options.

3.3.3 Double slit

Random events A and B in terms of probability[12] are independent (2.15) when $P(A \cap B) = P(A)P(B)$, and are exclusive when they are disjoint, $A \cap B = \varnothing$. Since $P(\varnothing) = 0$ is true only for impossible events, for possible events of positive probabilities, the definition of "independence" states that independent events are not exclusive and, conversely, that dependent events are exclusive (Theorem 67). For example, independent events in a probability distribution add up to a unit sum, such as the probabilities ($0 \le p, q \le 1$) of a (fair) coin landing on a "tail"or "head", whose sum is one ($p + q = 1$).

From the Kolmogorov probability axiom (56th definition) follows an interesting consequence, vi. $P(A) + P(B) = P(A \cup B) + P(A \cap B)$, which says that the sum of the probabilities of two events can be written as the sum of the probabilities of their union and intersection. Then the exclusivity $A \cap B = \varnothing$ means $P(A \cup B) =$

[12] "Independence" for probabilities does not have the same meaning as in "linear independence".

$P(A)+P(B)$, and from the previous $P(A\cap B)=P(A)P(B)$ if at least one of the events is impossible (empty set, $\varnothing$). With that knowledge, let's consider:

$$P(A\cup B)=P(A)+P(B)-P(A\cap B),$$

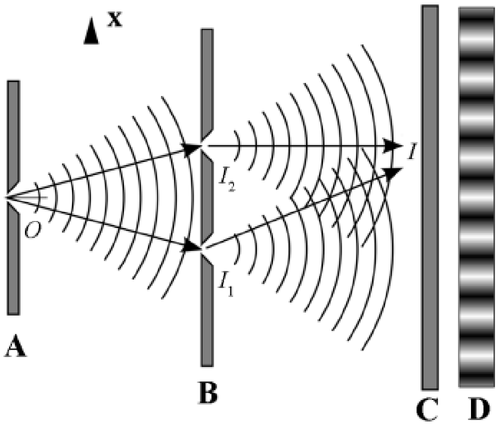
$$P(I)=P(I_1\cup I_2)=P(I_1)+P(I_2)-P(I_1\cap I_2),$$

$$|A|^2=|A_1+A_2|^2=(A_1+A_2)^*(A_1+A_2)=A_1^*A_1+A_1^*A_2+A_2A_1+A_2^*A_2,$$

$$|A|^2=|A_1|^2+2\Re(A_1^*A_2)+|A_2|^2, \tag{3.40}$$

where $P(I)=|A|^2$, $P(I_1)=|A_1|^2$, $P(I_2)=|A_2|^2$, $P(I_1\cap I_2)=2\Re(A_1^*A_2)$.

Here $A_1 = A(I_1)$ and $A_2 = A(I_2)$ are the amplitudes of the particle-waves passing through the slits I_1 and I_2 on the barrier B in the experiment double slit, in the figure on the right, and $A=A(I)$ is the amplitude of the output at the screen C. The waves exit the slit O, interfere, and form a diffraction image D ([10], p. 62) on the detection plate C. This is in accordance with classical physics.



However, when one of the slits, say I_1, is closed, then the "waves" behave like "particles" by grouping together on the line $O-I_2-I$. Even more strangely, no matter how scientists place detectors on the slits to detect which slit the particle-wave passes through, the D interference pattern disappears. This suggests that the very act of observing a photon "collapses" these many realities into one. For today's science, this is still incredible, but not for (this) information theory.

Namely, by declaring, placing detectors on the slit, or in any other way that would take away information from the wave, it would increase its certainty and form it as a particle. This is the same phenomenon as "by measuring an electron we define its previous path," or when, by opening a box with Schrödinger's cat, we find out whether it is alive or dead, which also creates a more certain past of the box in which the cat was not or was killed. This explanation is still undiscovered in the physics of this "strange" experiment, but it is quite consistent with the result (3.40).

Regarding the above "independence" versus "exclusivity," the case $I_1\cap I_2\ne\varnothing$ means that both openings are open and we have no information about which way the particle-wave goes. Then the equality $P(I_1\cap I_2)=P(I_1)P(I_2)=2\Re(A_1^*A_2)\ne 0$ holds, and there is interference on the path, i.e., diffraction at the target. Note that then there is no knowledge of the individual probabilities $P(I_1)$ and $P(I_2)$, so the uncertainty of the path remains until the end. On the contrary, when there is information about the probabilities $P(I_1)$ and $P(I_2)$, taking into account possible[13]

[13]Necessary assumption of Theorem 67.

events but also the known passage that did not occur, an either-or situation arises on the openings with $I_1 \cap I_2 = \varnothing$, from which it follows that $P(I_1 \cap I_2) \neq P(I_1)P(I_2)$, because the left side of the inequality is zero and the right side is not.

Another formalism is possible to describe this bizarre experiment. It is specific to this information theory, so I present it in a simplified form using a pair of complex numbers and their conjugates:

$$z_1 = x_1 + iy_1, \quad z_2 = x_2 + iy_2,$$

$$z_1^* = x_1 - iy_1, \quad z_2^* = x_2 - iy_2,$$

$$z_1^* z_2 = (x_1x_2 + y_1y_2) + i(x_1y_2 - x_2y_1),$$

$$z_1^* z_2 = \langle z_1, z_2 \rangle + i[z_1, z_2],$$

where $x_1, y_1, x_2, y_2 \in \mathbb{R}$, $i^2 = -1$. Compare this with the amplitudes above (3.40):

$$A_1^* A_2 = \langle A_1, A_2 \rangle + i[A_1, A_2], \tag{3.41}$$

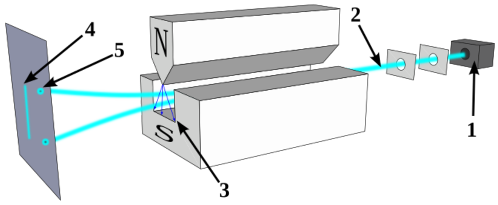
written symbolically to make it easier to see the interference, superposition, or internal coupling of information in the inner product $\langle A_1, A_2 \rangle$, and the external interference in the commutator, in the intensity of the outer product $[A_1, A_2]$. When there is no external communication, $[A_1, A_2] = 0$, which means that there is no knowledge of the aperture through which the wave passes, then the maximum internal coupling is $|\langle A_1, A_2 \rangle| = |A_1^* A_2|$. This is the superposition, or interference and diffraction of the waves. When there is external intervention (as in the case of central forces on a charge), then there is no diffraction at all, because at the microcosm level it is an "all or nothing" situation, it is quantized.

The case $I_1 \cap I_2 \neq \varnothing$ when both openings are open and there is no information about which path the particle-wave takes is $[A_1, A_2] = 0$, with no external reflection and the wave is completely introverted. The wave is extroverted, more loosely speaking, when $I_1 \cap I_2 = \varnothing$ but there is one used and one unused possibility, and then the commutator is maximal, $|[A_1, A_2]| = |A_1^* A_2|$.

3.3.4 Spin

In 1920, Otto Stern and Walter Gerlach devised an experiment that inadvertently led to the discovery of electron spin. They did this by placing silver foil in an oven to evaporate its atoms, and the silver atoms collected in a beam that passed through an inhomogeneous magnetic field. However, the magnet split the beam into two (and only two) separate parts.

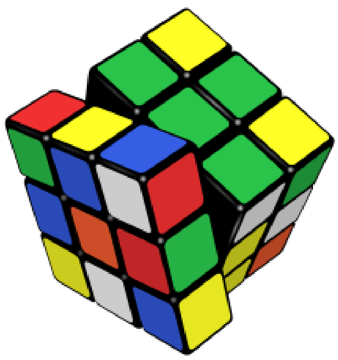
The following image on the left is the Stern-Gerlach experiment: Silver atoms travel through an inhomogeneous magnetic field and are deflected up or down depending on their spin: (1) furnace, (2) beam of silver atoms, (3) inhomogeneous magnetic field, (4) classically expected result, and (5) actual result. In quantum physics, this experiment showed that the spatial orientation of angular momentum is quantized.



It was first thought that electrons rotate, creating a magnetic field that causes these deflections, but this idea was soon rejected because calculations showed that this would require the electron to rotate faster than light. Today, particles (including the electron with mass $m = 0.511$ MeV/c^2, charge $q = -1\ e$, and spin $s = \frac{1}{2}$) are viewed as dimensionless points, and spin is treated as a permanent internal property of them, like mass or charge, but through the geometry of rotations, or group theory.

In mathematics, a *group* is a set G with an operation that satisfies closure (from $a, b \in G$ it follows $a \circ b \in G$), associativity ($(a \circ b) \circ c = a \circ (b \circ c)$, for all $a, b, c \in G$), must have a neutral element (there exists $e \in G$ such that $a \circ e = e \circ a = a$ for all $a \in G$), and every element must have a corresponding inverse element (every $a \in G$ has $a^{-1} \in G$ such that $a \circ a^{-1} = a^{-1} \circ a = e$).

Rubik's Cube (1974), pictured at right, is a common illustration of permutation groups. The elements of the Rubik's Cube group are rotations of a single horizontal or vertical facet. The 3×3×3 Rubik's Cube consists of 6 faces, each with 9 colored facelets, for a total of 54. Moving the cube rotates one of the 6 faces by $90°$, $180°$, or $-90°$. The central facet rotates around its axis but otherwise remains in the same position. A solved cube has all the facelets on each face in the same color. The moves of the cube are most often described with Singmaster's (David Breyer Singmaster, 1938-2023) notation.



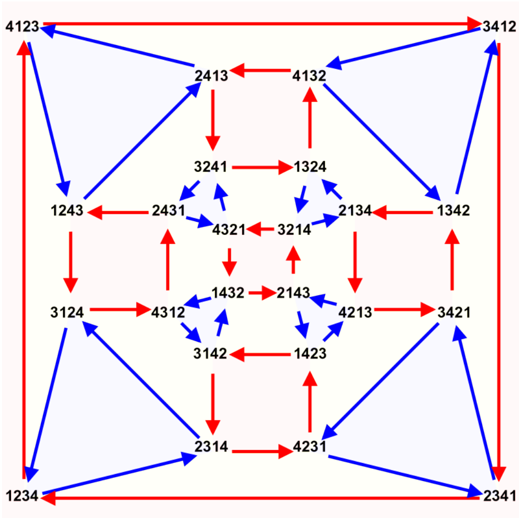
Permutation groups are ordered sequences of n-tuples with the operation of swapping elements. For example, the permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 5 & 4 & 3 & 1 \end{pmatrix}$$

is an operation that translates the upper numbers into the lower numbers, such that $\sigma(1) = 2$, $\sigma(2) = 5$, $\sigma(3) = 4$, $\sigma(4) = 3$, and $\sigma(5) = 1$. When applied twice to the original sequence, we obtain

$$\sigma^2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 1 & 3 & 4 & 2 \end{pmatrix}.$$

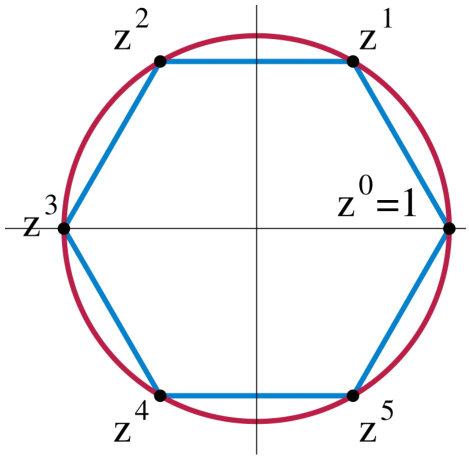
We can also write this as $\sigma^3(12345) = 12435$, when applied three times. Shorter written as $\sigma^4 = 24341$ after four consecutive applications, assuming the initial string is 12345. We then find $\sigma^5 = 51432$ and $\sigma^6 = 12345$. Although a five-element string has 5! = 120 permutations, the group generated by this binary operation has only six elements: $\{\sigma, \sigma^2, \sigma^3, \sigma^4, \sigma^5, \sigma^6\}$. A symmetric group S_n consists of all $n!$ permutations of some set of n elements.



The image on the left shows the symmetric group S_4 with all $4! = 24$ permutations of the string 1234. It is generated with two functions, the red arrow $f(abcd) = dabc$, and the blue arrow $g(abcd) = bcad$. The red ones (f) only give four permutations, and the blue ones (g) only three, but by combining them all 24 can be obtained.

A group with three elements that reduces to the initial element in three steps is denoted by Z_3, and is formed by this function g. A group that reaches the initial element in four steps is Z_4, which would be formed by this function f. The above function σ produces a group of the label Z_6. We often form and analyze them using rotations.

The rotations of Cartesian rectangular systems (3.28) for the angle $\phi = 2\pi/n$ form the group Z_n. For example, the rotations for $90^\circ = \pi/2$ are the group Z_4, and the rotations with the angle $\phi = 60^\circ$ are the group Z_6. We can also demonstrate this in the complex plane $\mathbb{C}$.



The image on the right is of the unit circle with degrees of the complex number $z = e^{i\pi/3}$, where $z^2 = e^{2i\pi/3}, ..., z^5 = e^{5i\pi/3}, z^6 = z^0 = 1$. This is again the group Z_6. From the different ways of forming these groups, we extract common properties, and this is the "group theory." We denote the groups of rotations in two dimensions by SO(2). Analogously, SO(3) denotes the groups of rotations in three dimensions, and in special relativity we have the groups SO(3, 1) of rotations in three dimensions of space and one of time.

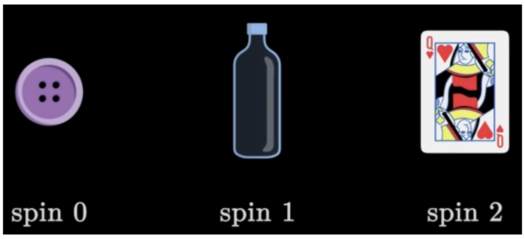
There are also infinite groups, say Z^2, such as an 8×8 chessboard whose only two rows, i.e., columns of black and white squares, would constantly alternate by shifting up or down, i.e., left or right.

In the following image, we have three types of gemstones. The first, blue, does not change at all when rotated by 90°. The second, red, requires four such rotations to return to its original appearance, and the third, green, requires two.

We do not know exactly what the particles themselves look like, but we can study them in an abstract state space and look for their geometric symmetries (Geometric explanation). We isolate those that are possibly analogous and very consistent with their physical behavior, and so, for example, we find that a blue stone would have spin zero ($s = 0$), a red one ($s = 1$), and a green two ($s = 2$). The spin number tells us how many rotations in physical space occur in abstract space. In

four steps, red repeats once, and green repeats twice.

During one full rotation, for 360° we calculate that the button in the next picture on the left does not have to be repeated at all ($s = 0$), that the bottle will be repeated once ($s = 1$), and the queen-heart card twice ($s = 2$). The spin of an object tells us how quickly it returns to its initial state. In physics, spin zero tells us that the object behaves as a scalar, as a number with no orientation in space. Spin one is a vector quantity, with orientations like the bottle in the picture. Spin two is a "double vector", or a tensor of order two, or a matrix. So much for the group Z_4.



For example, the Higgs boson (1964) is an elementary particle with spin $s = 0$, meaning it has no orientation in space. A photon is a particle with spin $s = 1$ that has both left and right orientations. Gravitons have spin $s = 2$, because general relativity views the gravitational field as second-order tensors. The half-integer spin of electrons and *fermions* $s = \frac{1}{2}$ then means that it takes two full spins (720°) for them to return to their initial state.

Spin was a latecomer to quantum mechanics. Even after the discovery of the now famous Schrödinger equation (1926), no one noticed that spin existed. That certain particles appear to be spinning, but they are not. Unlike general relativity, quantum mechanics and special relativity work very well together. For example, when an electron is measured, it is found that it can be in a spin-up $|\uparrow\rangle$ (clockwise) and a spin-down $|\downarrow\rangle$ (counterclockwise) state. An unobservable quantum system is in a superposition of these two states, say

$$|\phi\rangle = \frac{\sqrt{3}}{2}|\uparrow\rangle + \frac{1}{2}|\downarrow\rangle,$$

which means that the probability of finding the up state by measurement is $|\sqrt{3}/2|^2 = 3/4$, and of finding the spin-down state is $1/4$.

For a particle of mass m, electric charge q, with magnetic vector potential $\mathbf{A}$ and electric scalar potential ϕ, the *Pauli equation* is

$$\left[\frac{1}{2m}\left(\vec{\sigma}\cdot(\hat{\mathbf{p}} - q\mathbf{A})\right)^2 + q\phi\right]|\psi\rangle = i\hbar\frac{\partial}{\partial t}|\phi\rangle. \tag{3.42}$$

Here $\vec{\sigma} = (\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z)$ is the vector *Pauli operators*, represented by (2.4) *Pauli matri-*

ces:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and $\hat{\mathbf{p}} = -i\hbar\nabla = -i\hbar(\partial_x, \partial_y, \partial_z)$ is the momentum operator vector in the position representation. The wave function is:

$$|\phi\rangle = \phi_+|\uparrow\rangle + \phi_-|\downarrow\rangle = \begin{pmatrix} \phi_+ \\ \phi_- \end{pmatrix}.$$

The sum (probability) of the squares of the modulus is $|\phi_+|^2 + |\phi_-|^2 = 1$. The *Hamiltonian operator* is now a 2×2 matrix due to the Pauli operators

$$\hat{H} = \frac{1}{2m}[\vec{\sigma} \cdot (\hat{\mathbf{p}} - q\mathbf{A})]^2 + q\phi.$$

Putting it into the Schrödinger equation, it becomes the Pauli equation.

The spin is based on a mathematical model of symmetries (permutation groups), and if it corresponds to the observation of the Stern-Gerlach experiment, it means that it already meets two of the three criteria for "reality": (1) observability and (3) truth. Criterion (2) of the sustainability of reality follows from Noether's theorem, which states that every continuous symmetry of the interaction of a physical system with conservative forces has a corresponding conservation law. Therefore, the conservation law also applies to spin.

Forms of energy are fictions, and electron trajectories are the past (unless they are absent, if they are unexplained), but the amount of energy, the charge of the electron, and its spin are realities. The past has a limited duration and a quasi (half) type of reality, which should otherwise be considered very layered and at least as diverse as there are subjects of perception.

However, quantized quantities such as energy, charge, or spin are not elementary information but rather their indivisible packets into physical particles. Let us note that information theory also looks at the "interior" of an action, the quanta, but what it eventually "sees" there is not considered a topic of research in physics. In this way, it also understands the interior of, say, position or energy, but it does not see any physical, material objects there.

3.3.5 Entanglement

Quantum entanglement is the phenomenon of a group of particles that arise, interact with each other, or share spatial proximity in such a way that the quantum state of each of the group cannot be described independently of the states of the others, including when the particles are separated by a large distance.

This phenomenon was first observed as the *EPR paradox* (1935) when Einstein, together with Podolsky and Rosen, noticed "phantom action at a distance"while considering the theory of quantum mechanics and therefore doubted its algebraic foundations (Coupling). It can observe the same system in two spatially very distant ways, so that light cannot reach from the first part (A) to the second (B),

but the conservation law still manages to reconcile them. Einstein believed that there are some *hidden parameters* behind the formulas we know, which need to be discovered and this absurdity corrected.

In that joint work ([1], 1935), they disputed the well-known measurement expressions:

$$\Psi(x_1,x_2)=\sum_{n=1}^{\infty}\psi_n(x_2)u_n(x_1),\quad \Psi(x_1,x_2)=\sum_{s=1}^{\infty}\varphi_s(x_2)v_s(x_1). \tag{3.43}$$

The two equalities represent the same wave function $\Psi(x_1,x_2)$ in two ways allowed in quantum mechanics. In the first, there are measurements of some observable A of eigenvalues $a_1,a_2,a_3,\dots$ and their corresponding eigenfunctions $u_1(x_1)$, $u_2(x_1)$, $u_3(x_1)$, ..., where the variables x_1 describe the first system. The functions $\psi_n(x_2)$ of the coefficients of the linear combination, the expansion Ψ into an orthogonal sequence of functions $u_n(x_1)$. The assumption is that we measure A and that its value a_k has been found.

In the second, we choose B eigenvalues $b_1,b_2,b_3,\dots$ whose corresponding eigenfunctions are $v_1(x_1),v_2(x_1),v_3(x_1),\dots$, with $\varphi_s(x_2)$ new expansion coefficients. This B was measured, and the value of b_r was found immediately after the first system remained in state A. Therefore, it is possible to associate two different wave functions, ψ_k and φ_r, with the same reality, the authors conclude, considering this incorrect because it could happen that the measurement results are intertwined while systems A and B are too far apart for the second to react to changes in the first.

For an unusually long time, no one dealt with the problem of the EPR paradox until Bell ([2], 1964) proved the untenability of the idea of a solution with hidden parameters. To understand Bell's theorem, you do not need a particularly extensive knowledge of quantum mechanics, but rather a recognition of the impossibility of violating the probabilities he talks about there using the alleged parameters. Listen to the explanation, for example, in the video Bell's Inequality, or see it even more simply in the attachment Spooky Action. I will recount one point of the reduction of Bell's proof.

Suppose, for example, that one team of scientists rotates its electron spin detector relative to another lab by 180 degrees. This is equivalent to swapping its north and south poles so that an "up" result for one particle would never be followed by a "down" result for another. The scientists could also choose other rotations. Depending on the orientation of the magnets in the two labs, the probability of opposite results can range from 0% to 100%.

We interpret two measurements (3.43) in two laboratories (LAB 1, LAB 2) without going into the specific orientations of the three possible measurement axes marked A, B, and C in the figure below right. We observe pairs of electrons with spin "up" (green) and "down" (orange), and for each electron pair, each laboratory measures the spin of one of the electrons along one of these three axes chosen at random. Three such results can be classified into eight possibilities in the following figure.

If the world is described by a local hidden variable theory, not by quantum physics, each electron has its own spin value in each of three directions. For any electron pair that has spin values labeled 1 or 8, measurements in two laboratories will always give opposite results, no matter which axes the scientists choose to measure. The other six sets of spin values give opposite results in 33% of measurements on different axes. For example, for spin values labeled 5, the laboratories will get opposite outcomes when one measures along the B axis while the other measures along the C axis; this represents one-third of the possible choices.

Thus, laboratories will obtain opposite results when measuring along different axes at least 33% of the time; equivalently, they will obtain the same result at most 67% of the time. This result is the upper bound on the correlations allowed by local hidden variable theories, and it follows from Bell's inequality theorem. According to formula (3.43), when all three axes are spaced as far apart as possible, that is, all 120 degrees apart (as in the Mercedes logo), both laboratories will obtain the same result 75% of the time, which is too much for Bell's upper bound of 67%.

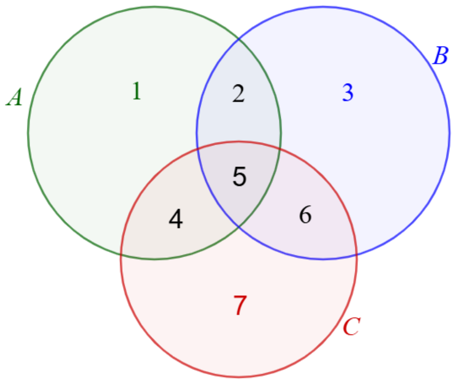
We will not go into the details of this calculation here but simply state the gist of Bell's theorem: if locality holds and a measurement of one particle will not instantaneously affect the outcome of another measurement far away, then the results in a given experimental setup have no correlation greater than 67%. If, on the other hand, the fates of entangled particles are inextricably linked even over vast distances, as predicted by the EPR paradox in quantum mechanics, the results of certain measurements will show stronger correlations.

Among the first experiments considered to be evidence of quantum entanglement were the 1969 proposals and 1972 performances by John Clauser and his colleagues. For this, they were awarded the 2010 Wolf Prize in Physics, awarded annually by the Wolf Foundation in Israel, and in 2022 they will also be the winners of the Nobel Prizes in Physics. Their findings are based on Bell's theorem.

Quantum entanglement is a type of "superposition" with more than one state, for example, like flipping a coin before seeing the result. Whereas "entanglement" is a special type of superposition that involves two spatially separated locations. A photon that encounters a half-and-half divider could be so entangled. By splitting, the photon could be on path A, or it could be on path B. In this case, the superposition is between a photon on path A that is not on path B and the absence of a photon on path A but its presence on path B.

We would think that it's just one or the other, but we just don't know which—and we'd be wrong. A photon is actually in both states until we measure it. Namely, the assumption that a particle has this imaginary definitive characteristic before it's

actually measured leads us into contradictions later. Why?



So we see it in the picture on the left, otherwise one of the explanations of Bell's inequality. In this interpretation, it says that "being in A and not being in B plus being in B and not being in C gives more than being in A and not being in C."Or in short: $A \wedge \neg B + B \wedge \neg C \geq A \wedge \neg C$. The picture shows a Venn diagram of sets, where parts of the sets A, B, and C are marked, say, people who know English, French, and German, in that order. The labels 1 are those who know only English, the labels 2 are those who know English and French but not German, ..., 5 who know all three languages, and 7 who know only German. The overlap is in the sentence "Knowledge of English without French (1, 4) plus French without German (2, 3) has more candidates than knowledge of English without German (1, 2)."

An example of quantum entanglement is a light source that emits two photons simultaneously. This pair of photons can be entangled so that the polarizations of both are of any orientation (i.e., random), but the two are always the same, or in a different experiment they are always opposite. The uncertainty lies in the measurement of the first, no matter which of the two, but it is quite certain what will be found by measuring the second, no matter how spatially distant it is from the first at that moment. This enigmatic entanglement has already been confirmed by enough experiments that physics no longer doubts it, but alas, it is still without a reasonable explanation.

However, my information theory has such a simple and self-consistent interpretation of this quantum entanglement that I always feel sorry to explain it and spoil that beautiful magic in physics. Nevertheless, I have brought it up several times privately in the blog and in conversations as a side topic to some "more serious"observations.

If I were to interpret “quantum entanglement” using “information theory” (mine), I would do so using the concept of “present.” The states of reality are vectors, from the few quantum ones upwards, that develop in time and always around some isometries, their own backbones. Themselves, what is being mirrored, are presents, images of information that are relative (1.3.5 Simultaneity), not only dependent on movement but on subjects in general, among whom there are no two that see the surrounding world in the same way (2.4.2 Uniqueness). Hence, the true properties of “present” are difficult to discern.

Glava 4

Conclusion

Teaching science is not science, and neither is private opinion about science in the manner of science. My blogs are something in between, too intimate to share publicly and too personal to selfishly keep. These excursions into scripts about reflections that are someone's occupations are a kind of hobby and need for freedom of expression, but above all in matters of ultimate truths. And they turned out to be narrow essays or studies, let's say without reference to the transitions from inanimate to living matter, which for many could be an interesting topic for my own brainstorming, reflections on the rise and fall of societies, democracies, or, for example, a slightly more extensive story about the well-known formalisms of information.

Slika 4.1: Grandfather and grandson, November 5, 2024. Banja Luka.

Bibliografija

[1] Einstein, A.; Podolsky, B.; Rosen, N. (1935-05-15). "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?"

[2] Bell, J. S. (1964). "On the Einstein Podolsky Rosen Paradox". Physics Physique Физика.

[3] S.Aljančić: UVOD U REALNU I FUNKCIONALNU ANALIZU; Građevinska knjiga, Beograd, 1974. наше

[4] Zoran A. Ivković: TEORIJA VEROVATNOĆA SA MATEMATIČKOM STATISTIKOM; Građevinska knjiga, Beograd, 1976.

[5] Rastko Vuković: Konusni preseci; Arhimed, Banja Luka 2014.

[6] Rastko Vuković: MATEMATIČKA TEORIJA INFORMACIJE I KOMUNIKACIJE; Društvo matematičara Republike Srpske, Banja Luka, 1995.

[7] Rastko Vuković: INFORMATION of PERCEPTION – freedom, democracy and physics; Економски институт Бања Лука, 2016.

[8] Rastko Vuković: SPACE-TIME – principles physics of chances; Economic Institute Banja Luka, 2017.

[9] Rastko Vuković: MULTIPLICITIES – Choices for animate and inanimate; Economic Institute Banja Luka, 2018.

[10] Rastko Vuković: QUANTUM MECHANICS – representation of the continuum. After reviews in Banja Luka in 2019, the book is in digital format.

[11] Rastko Vuković: PHYSICAL INFORMATION – in social phenomena and mathematics; Economic Institute Banja Luka, 2019.

[12] Rastko Vuković: MINIMALISM of INFORMATION – physical information and applications; Economic Institute Banja Luka, 2020.

[13] Rastko Vuković: Informatička Teorija – Part 1, contributions to private conversations, May-July 2022.

Indeks